



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 125446

TO: Manjunath N Rao
Location: rem/3b81/3c70
Art Unit: 1652
Monday, June 28, 2004

Case Serial Number: 10/038723

From: Alex Waclawiw
Location: Biotech-Chem Library
Rem 1A71
Phone: 272-2534

Alexandra.waclawiw@uspto.gov

Search Notes

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:44:58 ; Search time 50 Seconds
(without alignments)
3015.110 Million cell updates/sec

Title: US-10-038-723-2
Perfect score: 2771
Sequence: 1 MSFRLSLGLSVCTGLANV.....SKTTATASKSTTTTBSGMSL 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2771	100.0	534	9	US-09-808-395-2
2	2771	100.0	534	10	US-09-821-616-9
3	2771	100.0	534	14	US-10-038-723-2
4	2771	100.0	534	15	US-10-421-586-2
5	2771	100.0	534	15	US-10-421-586-3
6	2742	99.0	640	14	US-10-038-723-13
7	2742	99.0	640	15	US-10-421-586-14
8	2735	98.7	743	15	US-10-418-836-9
9	2735	98.7	743	15	US-10-418-836-13
10	2735	98.7	743	15	US-10-418-836-10
11	2735	98.7	743	15	US-10-418-836-16
12	2720.5	98.2	748	15	US-10-418-836-19
13	2716	98.0	741	15	US-10-418-836-30
14	2715.5	98.0	972	15	US-10-418-836-38
15	2715	98.0	739	15	US-10-418-836-26

16	2715	98.0	739	15	US-10-418-836-27	Sequence 27, Appl
17	2714.5	98.0	740	15	US-10-418-836-28	Sequence 28, Appl
18	2714.5	98.0	740	15	US-10-418-836-34	Sequence 34, Appl
19	2714.5	98.0	742	15	US-10-418-836-29	Sequence 29, Appl
20	2714	97.9	743	15	US-10-418-836-31	Sequence 31, Appl
21	2714	97.9	743	15	US-10-418-836-35	Sequence 35, Appl
22	2714	97.9	743	15	US-10-418-836-39	Sequence 39, Appl
23	2523	91.1	1095	15	US-10-228-063-45	Sequence 45, Appl
24	1850.5	66.8	631	14	US-10-213-990-54	Sequence 34, Appl
25	1689	61.0	618	10	US-09-821-616-34	Sequence 7, Appl
26	1674.5	60.4	591	10	US-09-821-616-7	Sequence 7, Appl
27	1548.5	55.9	647	15	US-10-369-493-3109	Sequence 3109, Ap
28	1425.5	51.4	581	14	US-10-281-673-3	Sequence 3, Appl
29	1408.5	50.8	704	14	US-10-213-990-21	Sequence 21, Appl
30	1380.5	49.8	620	14	US-10-213-990-33	Sequence 33, Appl
31	1250	45.1	441	15	US-10-369-493-12402	Sequence 12402, A
32	1132	40.9	432	15	US-10-369-493-12533	Sequence 12533, A
33	781.5	28.2	450	15	US-10-369-493-2490	Sequence 2490, Ap
34	725	26.2	579	14	US-10-228-063-49	Sequence 49, Appl
35	597	21.5	549	15	US-10-369-493-22096	Sequence 22096, A
36	164.5	5.9	2344	9	US-09-815-242-12713	Sequence 12713, A
37	160	5.8	35	10	US-09-821-616-4	Sequence 4, Appl
38	159.5	5.8	2283	14	US-10-172-502-4	Sequence 4, Appl
39	159	5.7	2271	12	US-10-282-122A-43924	Sequence 43924, A
40	158	5.7	1283	15	US-10-369-493-22616	Sequence 22616, A
41	155.5	5.6	1621	14	US-10-185-990-10	Sequence 10, Appl
42	155.5	5.6	1621	14	US-10-185-990-11	Sequence 11, Appl
43	154.5	5.6	1031	9	US-09-815-242-10932	Sequence 10932, A
44	152.5	5.5	45	15	US-10-120-801-63	Sequence 63, Appl
45	152	5.5	45	15	US-10-128-063-53	Sequence 53, Appl
46	151.5	5.5	1099	12	US-10-282-122A-48220	Sequence 48220, A
47	150	5.4	1099	15	US-10-369-493-1535	Sequence 1535, Ap
48	147.5	5.3	4440	12	US-10-206-915-525	Sequence 525, App
49	147.5	5.3	4440	12	US-10-199-670-525	Sequence 525, App
50	147.5	5.3	4440	12	US-10-201-858-525	Sequence 525, App
51	147.5	5.3	4440	12	US-10-205-890-525	Sequence 525, App
52	147.5	5.3	4440	12	US-10-208-024-525	Sequence 525, App
53	147.5	5.3	4440	12	US-10-201-853-525	Sequence 525, App
54	147.5	5.3	4440	12	US-10-174-581-525	Sequence 525, App
55	147.5	5.3	4440	12	US-10-176-483-525	Sequence 525, App
56	147.5	5.3	4440	12	US-10-176-749-525	Sequence 525, App
57	147.5	5.3	4440	12	US-10-176-914-525	Sequence 525, App
58	147.5	5.3	4440	12	US-10-176-915-525	Sequence 525, App
59	147.5	5.3	4440	12	US-10-176-915-525	Sequence 525, App
60	147.5	5.3	4440	12	US-10-180-550-525	Sequence 525, App
61	147.5	5.3	4440	12	US-10-183-014-525	Sequence 525, App
62	147.5	5.3	4440	12	US-10-187-718-525	Sequence 525, App
63	147.5	5.3	4440	12	US-10-187-740-525	Sequence 525, App
64	147.5	5.3	4440	12	US-10-187-883-525	Sequence 525, App
65	147.5	5.3	4440	12	US-10-194-353-525	Sequence 525, App
66	147.5	5.3	4440	12	US-10-194-460-525	Sequence 525, App
67	147.5	5.3	4440	12	US-10-194-463-525	Sequence 525, App
68	147.5	5.3	4440	12	US-10-194-484-525	Sequence 525, App
69	147.5	5.3	4440	12	US-10-195-884-525	Sequence 525, App
70	147.5	5.3	4440	12	US-10-195-896-525	Sequence 525, App
71	147.5	5.3	4440	12	US-10-196-744-525	Sequence 525, App
72	147.5	5.3	4440	12	US-10-196-755-525	Sequence 525, App
73	147.5	5.3	4440	12	US-10-196-757-525	Sequence 525, App
74	147.5	5.3	4440	12	US-10-197-704-525	Sequence 525, App
75	147.5	5.3	4440	12	US-10-197-710-525	Sequence 525, App
76	147.5	5.3	4440	12	US-10-198-758-525	Sequence 525, App
77	147.5	5.3	4440	12	US-10-198-766-525	Sequence 525, App
78	147.5	5.3	4440	12	US-10-199-304-525	Sequence 525, App
79	147.5	5.3	4440	12	US-10-199-309-525	Sequence 525, App
80	147.5	5.3	4440	12	US-10-199-313-525	Sequence 525, App
81	147.5	5.3	4440	12	US-10-199-456-525	Sequence 525, App
82	147.5	5.3	4440	12	US-10-201-329-525	Sequence 525, App
83	147.5	5.3	4440	12	US-10-202-412-525	Sequence 525, App
84	147.5	5.3	4440	12	US-10-206-919-525	Sequence 525, App
85	147.5	5.3	4440	12	US-10-206-922-525	Sequence 525, App
86	147.5	5.3	4440	12	US-10-206-924-525	Sequence 525, App
87	147.5	5.3	4440	12	US-10-206-928-525	Sequence 525, App
88	147.5	5.3	4440	12	US-10-207-914-525	Sequence 525, App

```
89 147.5 5.3 4440 12 US-10-207-921-525 Sequence 525, App
90 147.5 5.3 4440 12 US-10-207-922-525 Sequence 525, App
91 147.5 5.3 4440 12 US-10-208-027-525 Sequence 525, App
92 147.5 5.3 4440 12 US-10-174-570-525 Sequence 525, App
93 147.5 5.3 4440 12 US-10-183-005-525 Sequence 525, App
94 147.5 5.3 4440 13 US-10-052-586-525 Sequence 525, App
95 147.5 5.3 4440 14 US-10-174-590-525 Sequence 525, App
96 147.5 5.3 4440 14 US-10-176-758-525 Sequence 525, App
97 147.5 5.3 4440 14 US-10-175-737-525 Sequence 525, App
98 147.5 5.3 4440 14 US-10-173-706-525 Sequence 525, App
99 147.5 5.3 4440 14 US-10-175-738-525 Sequence 525, App
100 147.5 5.3 4440 14 US-10-175-752-525 Sequence 525, App
```

ALIGNMENTS

```
RESULT 1
US-09-908-395-2
; Sequence 2, Application US/09908395
; Patent No. US20020164723A1
; GENERAL INFORMATION:
; APPLICANT: Liaw, Gin
; APPLICANT: Pedersen, Sven
; APPLICANT: Hendriksen, Sven
; TITLE OF INVENTION: A Method of Producing Saccharide
; FILE REFERENCE: 5318.200-US
; CURRENT APPLICATION NUMBER: US/09/908,395
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US/09/198,672
; PRIOR FILING DATE: PRIOR FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-908-395-2
```

```
Query Match 100.0%; Score 2771; DB 9; Length 534;
Best Local Similarity 100.0%; Pred. No. 2.7e-237;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRSLALSGLVCTGTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
DB 1 MSFRSLALSGLVCTGTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
QY 61 VVASPSTNDPDIYTYTWRDSGLVKTLYVDFRNGDTLSLSTIENYISAQAIYVGISNPSG 120
DB 61 VVASPSTNDPDIYTYTWRDSGLVKTLYVDFRNGDTLSLSTIENYISAQAIYVGISNPSG 120
QY 121 DLSGAGLGPKNVDETAATYATGSGWRPQRDGPALRATAMIFGQWMLDNGYSTATDIW 180
DB 121 DLSGAGLGPKNVDETAATYATGSGWRPQRDGPALRATAMIFGQWMLDNGYSTATDIW 180
QY 121 DLSGAGLGPKNVDETAATYATGSGWRPQRDGPALRATAMIFGQWMLDNGYSTATDIW 180
DB 121 DLSGAGLGPKNVDETAATYATGSGWRPQRDGPALRATAMIFGQWMLDNGYSTATDIW 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240
QY 241 APRIILCYLQSFWMGFTILANFDSRSRGKDANTLLGSIHTFPDPAACDSTFPQCSPRALA 300
DB 241 APRIILCYLQSFWMGFTILANFDSRSRGKDANTLLGSIHTFPDPAACDSTFPQCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEDTYNGNPMWFLCTTLAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEDTYNGNPMWFLCTTLAAEQLYDALYQMD 360
QY 361 KQGSLEVTDVSLDFFKALYSDAATGYSSSSSYTSIVDAVKTFADGPFVSI VETHAASNG 420
DB 361 KQGSLEVTDVSLDFFKALYSDAATGYSSSSSYTSIVDAVKTFADGPFVSI VETHAASNG 420
```

```
DB 361 KQGSLEVTDVSLDFFKALYSDAATGYSSSSSYTSIVDAVKTFADGPFVSI VETHAASNG 420
QY 421 SMSQYDKSDGDELSARDLWTSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSQYDKSDGDELSARDLWTSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TVSSVTVTSWPSIVATGCTTTTATPTPGSGSVTSTKTATATSTTTTSGMGL 534
DB 481 TVSSVTVTSWPSIVATGCTTTTATPTPGSGSVTSTKTATATSTTTTSGMGL 534
```

```
RESULT 2
US-09-821-616-9
; Sequence 9, Application US/09821616
; Publication No. US20030027290A1
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne R.
; APPLICANT: Nielsen, Ruby
; APPLICANT: Lehmebeck, Dan
; TITLE OF INVENTION: Thermostable Glucoamylase
; FILE REFERENCE: 5279.200-US
; CURRENT APPLICATION NUMBER: US/09/821,616
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/199,290
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1557/97
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-30
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 0925/98
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,746
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/094,344
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/979,673
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/107,657
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-821-616-9
```

```
Query Match 100.0%; Score 2771; DB 10; Length 534;
Best Local Similarity 100.0%; Pred. No. 2.7e-237;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRSLALSGLVCTGTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
DB 1 MSFRSLALSGLVCTGTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
QY 61 VVASPSTNDPDIYTYTWRDSGLVKTLYVDFRNGDTLSLSTIENYISAQAIYVGISNPSG 120
DB 61 VVASPSTNDPDIYTYTWRDSGLVKTLYVDFRNGDTLSLSTIENYISAQAIYVGISNPSG 120
QY 121 DLSGAGLGPKNVDETAATYATGSGWRPQRDGPALRATAMIFGQWMLDNGYSTATDIW 180
DB 121 DLSGAGLGPKNVDETAATYATGSGWRPQRDGPALRATAMIFGQWMLDNGYSTATDIW 180
QY 121 DLSGAGLGPKNVDETAATYATGSGWRPQRDGPALRATAMIFGQWMLDNGYSTATDIW 180
DB 121 DLSGAGLGPKNVDETAATYATGSGWRPQRDGPALRATAMIFGQWMLDNGYSTATDIW 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240
QY 241 APRIILCYLQSFWMGFTILANFDSRSRGKDANTLLGSIHTFPDPAACDSTFPQCSPRALA 300
DB 241 APRIILCYLQSFWMGFTILANFDSRSRGKDANTLLGSIHTFPDPAACDSTFPQCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEDTYNGNPMWFLCTTLAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEDTYNGNPMWFLCTTLAAEQLYDALYQMD 360
```



```
Db 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPMFLCTIAAEQYDALYQWD 360
Qy 361 KQGSLEVTVDLDFPFAALYSDAATGYSSSSSTYSSIVDAVKTFADGPFVSIYETHAASNG 420
Db 361 KQGSLEVTVDLDFPFAALYSDAATGYSSSSSTYSSIVDAVKTFADGPFVSIYETHAASNG 420
Qy 421 SMSQYDKSDGDEQLSARDLTWGSYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
Db 421 SMSQYDKSDGDEQLSARDLTWGSYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
Qy 481 TYSSVTVTWPSPSVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTTSGMSL 534
Db 481 TYSSVTVTWPSPSVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTTSGMSL 534

RESULT 3
US-10-038-723-2
; Sequence 2, Application US/10038723
; Publication No. US20030032163A1
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne Roenfeldt
; APPLICANT: Svendsen, Allan
; APPLICANT: Pedersen, Henrik
; APPLICANT: Vind, Jesper
; APPLICANT: Hendriksen, Hanne Vang
; APPLICANT: Frandsen, Torben Peter
; TITLE OF INVENTION: Glucoamylase Variants
; FILE REFERENCE: 5636.200-US
; CURRENT APPLICATION NUMBER: US/10/038,723
; PRIOR FILING DATE: 2002-01-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/351,814
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PA 1998 00937
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PA 1998 01667
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/093,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/115,545
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-10-038-723-2

Query Match 100.0%; Score 2771; DB 14; Length 534;
Best Local Similarity 100.0%; Pred. No. 2,7e-237;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
Db 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
Qy 61 VVASPSTNDPDEYTYWTRDSGLVLTIVDLFRNGDTSLSLTIBNYISAQAIYVGISNPSG 120
Db 61 VVASPSTNDPDEYTYWTRDSGLVLTIVDLFRNGDTSLSLTIBNYISAQAIYVGISNPSG 120
Qy 121 DLSGAGLGEPPKNNVDETAATYTGSGRQPDGPAIRATAMIGFQWLLDNGYSTTADIW 180
Db 121 DLSGAGLGEPPKNNVDETAATYTGSGRQPDGPAIRATAMIGFQWLLDNGYSTTADIW 180
Qy 181 PLVRNDLSYVAQYWNQGTGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVAGSSCSWCDQ 240
Db 181 PLVRNDLSYVAQYWNQGTGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVAGSSCSWCDQ 240
Qy 241 APPEILCYLQSFMTGSFILANFDSRSRSGKDNLTLLGSIHTFDPEAACDDSTFQCSPRALA 300
```

```
Db 241 APPEILCYLQSFMTGSFILANFDSRSRSGKDNLTLLGSIHTFDPEAACDDSTFQCSPRALA 300
Qy 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPMFLCTIAAEQYDALYQWD 360
Db 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPMFLCTIAAEQYDALYQWD 360
Qy 361 KQGSLEVTVDLDFPFAALYSDAATGYSSSSSTYSSIVDAVKTFADGPFVSIYETHAASNG 420
Db 361 KQGSLEVTVDLDFPFAALYSDAATGYSSSSSTYSSIVDAVKTFADGPFVSIYETHAASNG 420
Qy 421 SMSQYDKSDGDEQLSARDLTWGSYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
Db 421 SMSQYDKSDGDEQLSARDLTWGSYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
Qy 481 TYSSVTVTWPSPSVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTTSGMSL 534
Db 481 TYSSVTVTWPSPSVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTTSGMSL 534

RESULT 4
US-10-421-586-2
; Sequence 2, Application US/10421586
; Publication No. US20040002142A1
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne Roenfeldt
; APPLICANT: Svendsen, Allan
; APPLICANT: Pedersen, Henrik
; APPLICANT: Vind, Jesper
; APPLICANT: Hendriksen, Hanne Vang
; APPLICANT: Frandsen, Torben Peter
; TITLE OF INVENTION: Glucoamylase Variants
; FILE REFERENCE: 5967.210-US
; CURRENT APPLICATION NUMBER: US/10/421,586
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-421-586-2

Query Match 100.0%; Score 2771; DB 15; Length 534;
Best Local Similarity 100.0%; Pred. No. 2,7e-237;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
Db 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
Qy 61 VVASPSTNDPDEYTYWTRDSGLVLTIVDLFRNGDTSLSLTIBNYISAQAIYVGISNPSG 120
Db 61 VVASPSTNDPDEYTYWTRDSGLVLTIVDLFRNGDTSLSLTIBNYISAQAIYVGISNPSG 120
Qy 121 DLSGAGLGEPPKNNVDETAATYTGSGRQPDGPAIRATAMIGFQWLLDNGYSTTADIW 180
Db 121 DLSGAGLGEPPKNNVDETAATYTGSGRQPDGPAIRATAMIGFQWLLDNGYSTTADIW 180
Qy 181 PLVRNDLSYVAQYWNQGTGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVAGSSCSWCDQ 240
Db 181 PLVRNDLSYVAQYWNQGTGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVAGSSCSWCDQ 240
Qy 241 APPEILCYLQSFMTGSFILANFDSRSRSGKDNLTLLGSIHTFDPEAACDDSTFQCSPRALA 300
Db 241 APPEILCYLQSFMTGSFILANFDSRSRSGKDNLTLLGSIHTFDPEAACDDSTFQCSPRALA 300
Qy 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPMFLCTIAAEQYDALYQWD 360
Db 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPMFLCTIAAEQYDALYQWD 360
Qy 361 KQGSLEVTVDLDFPFAALYSDAATGYSSSSSTYSSIVDAVKTFADGPFVSIYETHAASNG 420
```

Db 361 KQSGLEVTDVSLDPFKALYSADATGTYSSSSSTYSIVDAVKTPADGFSIVETHAASNG 420
Qy 421 SMSEQYDKSDGEOLASARDLTWMSYAALLTANNRRNSVVPASWGETSASSVPGTCAATSAIG 480
Db 421 SMSEQYDKSDGEOLASARDLTWMSYAALLTANNRRNSVVPASWGETSASSVPGTCAATSAIG 480
Qy 481 TYSSVTVTSPSIVATGCTTTTATPTGSGSVTSTSKTTATATASKTSTTRGMSL 534
Db 481 TYSSVTVTSPSIVATGCTTTTATPTGSGSVTSTSKTTATATASKTSTTRGMSL 534

RESULT 5

US-10-421-586-3
; Sequence 3, Application US/10421586
; Publication No. US20040002142A1
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne Ronfeldt
; APPLICANT: Svendsen, Allan
; APPLICANT: Pedersen, Henrik
; APPLICANT: Vind, Jesper
; APPLICANT: Hendriksen, Hanne Vang
; APPLICANT: Frandsen, Torben Peter
; TITLE OF INVENTION: Glucoamylase Variants
; FILE REFERENCE: 5967 210-US
; CURRENT APPLICATION NUMBER: US/10/421,586
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 534
; TYPE: PRF
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: Signal
; LOCATION: (1)..(24)
US-10-421-586-3

Query Match 100.0%; Score 2771; DB 15; Length 534;
Best Local Similarity 100.0%; Pred. No. 2,7e-237;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFRSLIALSGVCTGLANVTSKRATLDSWLSNEATYARFAILLNIGADGAWVSGADSGI 60
Db 1 MSFRSLIALSGVCTGLANVTSKRATLDSWLSNEATYARFAILLNIGADGAWVSGADSGI 60
Qy 61 VVASPSTNDPDIYFTYTRDSGLVKTLYDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
Db 61 VVASPSTNDPDIYFTYTRDSGLVKTLYDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
Qy 121 DLSGGAGLGPKNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYSTATDIW 180
Db 121 DLSGGAGLGPKNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYSTATDIW 180
Qy 181 PLYRNDLSYVAQYWNQGYDLMEEVNSSFPTTAVQHRALVEGSAFATAVGSSCSWCDQ 240
Db 181 PLYRNDLSYVAQYWNQGYDLMEEVNSSFPTTAVQHRALVEGSAFATAVGSSCSWCDQ 240
Qy 241 APETLCYLOSFWTGSFLLANFDSRSGKDNLTLLGSIHTFDPEAACDDSTFOPCSPRALA 300
Db 241 APETLCYLOSFWTGSFLLANFDSRSGKDNLTLLGSIHTFDPEAACDDSTFOPCSPRALA 300
Qy 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFLLCTLLAAEQLYDALYQWD 360
Db 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFLLCTLLAAEQLYDALYQWD 360
Qy 361 KQSGLEVTDVSLDPFKALYSADATGTYSSSSSTYSIVDAVKTPADGFSIVETHAASNG 420
Db 361 KQSGLEVTDVSLDPFKALYSADATGTYSSSSSTYSIVDAVKTPADGFSIVETHAASNG 420
Qy 421 SMSEQYDKSDGEOLASARDLTWMSYAALLTANNRRNSVVPASWGETSASSVPGTCAATSAIG 480
Db 421 SMSEQYDKSDGEOLASARDLTWMSYAALLTANNRRNSVVPASWGETSASSVPGTCAATSAIG 480

Qy 481 TYSSVTVTSPSIVATGCTTTTATPTGSGSVTSTSKTTATATASKTSTTRGMSL 534
Db 481 TYSSVTVTSPSIVATGCTTTTATPTGSGSVTSTSKTTATATASKTSTTRGMSL 534

RESULT 6

US-10-038-723-13
; Sequence 13, Application US/10038723
; Publication No. US20030032163A1
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne Ronfeldt
; APPLICANT: Svendsen, Allan
; APPLICANT: Pedersen, Henrik
; APPLICANT: Vind, Jesper
; APPLICANT: Hendriksen, Hanne Vang
; APPLICANT: Frandsen, Torben Peter
; TITLE OF INVENTION: Glucoamylase Variants
; FILE REFERENCE: 5636,200-US
; CURRENT APPLICATION NUMBER: US/10/038,723
; CURRENT FILING DATE: 2002-01-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/351,814
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PA 1998 00937
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PA 1998 01667
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/093,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/115,545
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 640
; TYPE: PRF
; ORGANISM: ASPERGILLUS NIGER
US-10-038-723-13

Query Match 99.0%; Score 2742; DB 14; Length 640;
Best Local Similarity 99.6%; Pred. No. 1,4e-234;
Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSFRSLIALSGVCTGLANVTSKRATLDSWLSNEATYARFAILLNIGADGAWVSGADSGI 60
Db 1 MSFRSLIALSGVCTGLANVTSKRATLDSWLSNEATYARFAILLNIGADGAWVSGADSGI 60
Qy 61 VVASPSTNDPDIYFTYTRDSGLVKTLYDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
Db 61 VVASPSTNDPDIYFTYTRDSGLVKTLYDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
Qy 121 DLSGGAGLGPKNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYSTATDIW 180
Db 121 DLSGGAGLGPKNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGYSTATDIW 180
Qy 181 PLYRNDLSYVAQYWNQGYDLMEEVNSSFPTTAVQHRALVEGSAFATAVGSSCSWCDQ 240
Db 181 PLYRNDLSYVAQYWNQGYDLMEEVNSSFPTTAVQHRALVEGSAFATAVGSSCSWCDQ 240
Qy 241 APETLCYLOSFWTGSFLLANFDSRSGKDNLTLLGSIHTFDPEAACDDSTFOPCSPRALA 300
Db 241 APETLCYLOSFWTGSFLLANFDSRSGKDNLTLLGSIHTFDPEAACDDSTFOPCSPRALA 300
Qy 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFLLCTLLAAEQLYDALYQWD 360
Db 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFLLCTLLAAEQLYDALYQWD 360
Qy 361 KQSGLEVTDVSLDPFKALYSADATGTYSSSSSTYSIVDAVKTPADGFSIVETHAASNG 420
Db 361 KQSGLEVTDVSLDPFKALYSADATGTYSSSSSTYSIVDAVKTPADGFSIVETHAASNG 420
Qy 421 SMSEQYDKSDGEOLASARDLTWMSYAALLTANNRRNSVVPASWGETSASSVPGTCAATSAIG 480
Db 421 SMSEQYDKSDGEOLASARDLTWMSYAALLTANNRRNSVVPASWGETSASSVPGTCAATSAIG 480

QY 481 TVSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTS 530
Db 481 TVSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTS 530

RESULT 7
US-10-421-586-14
; Sequence 14, Application US/10421586
; Publication No. US20040002142A1
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne Ronfeldt
; APPLICANT: Svendsen, Allan
; APPLICANT: Pedersen, Henrik
; APPLICANT: Vind, Jesper
; APPLICANT: Hendriksen, Hanne Vang
; APPLICANT: Frandsen, Torben Peter
; TITLE OF INVENTION: Glucoamylase Variants
; FILE REFERENCE: 5967.210-US
; CURRENT APPLICATION NUMBER: US/10/421,586
; CURRENT FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 640
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-421-586-14

Query Match 99.0%; Score 2742; DB 15; Length 640;
Best Local Similarity 99.6%; Pred. No. 1,4e-234;
Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSFRSLALSLGLVCTGLANVISKRATLDSWLSNEATVATATLNNNGADGAWVSGADSGI 60
Db 1 MSFRSLALSLGLVCTGLANVISKRATLDSWLSNEATVATATLNNNGADGAWVSGADSGI 60
QY 61 VVASPESTNDPDEYFTYTRDSGLVLTLDLFRNGDTSLSLTENYISAQAIYOGISNPSG 120
Db 61 VVASPESTNDPDEYFTYTRDSGLVLTLDLFRNGDTSLSLTENYISAQAIYOGISNPSG 120
QY 121 DLSGAGLCEPKFNVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIW 180
Db 121 DLSGAGLCEPKFNVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIW 180
QY 121 DLSGAGLCEPKFNVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIW 180
Db 121 DLSGAGLCEPKFNVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIW 180
QY 181 PLVRNDSLVAQYWNQTYDLMEEVNGSSFFITAVOHRALVEGSAFATVAGSSCWCDSQ 240
Db 181 PLVRNDSLVAQYWNQTYDLMEEVNGSSFFITAVOHRALVEGSAFATVAGSSCWCDSQ 240
QY 241 APEILCYLQSFMTGFTILANFDSRSRGKXANTLLGSIHTFDEPAACDDSTFQPCSPRALA 300
Db 241 APEILCYLQSFMTGFTILANFDSRSRGKXANTLLGSIHTFDEPAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAAEQLYDALYQMD 360
Db 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAAEQLYDALYQMD 360
QY 361 KQGSLEVTDSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTFADGFSIVETHAASNG 420
Db 361 KQGSLEVTDSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTFADGFSIVETHAASNG 420
QY 421 SMSEQYDKSDGEOLSRDLTWSYAALLTANRRNSVVPASWGETSASSVPGTCAATSAIG 480
Db 421 SMSEQYDKSDGEOLSRDLTWSYAALLTANRRNSVVPASWGETSASSVPGTCAATSAIG 480
QY 481 TVSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTS 530
Db 481 TVSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTS 530

RESULT 8
US-10-418-836-9
; Sequence 9, Application US/10418836

; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huang
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
; FILE REFERENCE: GCT41-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-9

Query Match 98.7%; Score 2735; DB 15; Length 743;
Best Local Similarity 99.6%; Pred. No. 7,2e-234;
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSFRSLALSLGLVCTGLANVISKRATLDSWLSNEATVATATLNNNGADGAWVSGADSGI 60
Db 1 MSFRSLALSLGLVCTGLANVISKRATLDSWLSNEATVATATLNNNGADGAWVSGADSGI 60
QY 61 VVASPESTNDPDEYFTYTRDSGLVLTLDLFRNGDTSLSLTENYISAQAIYOGISNPSG 120
Db 61 VVASPESTNDPDEYFTYTRDSGLVLTLDLFRNGDTSLSLTENYISAQAIYOGISNPSG 120
QY 121 DLSGAGLCEPKFNVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIW 180
Db 121 DLSGAGLCEPKFNVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIW 180
QY 121 DLSGAGLCEPKFNVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIW 180
Db 121 DLSGAGLCEPKFNVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIW 180
QY 181 PLVRNDSLVAQYWNQTYDLMEEVNGSSFFITAVOHRALVEGSAFATVAGSSCWCDSQ 240
Db 181 PLVRNDSLVAQYWNQTYDLMEEVNGSSFFITAVOHRALVEGSAFATVAGSSCWCDSQ 240
QY 241 APEILCYLQSFMTGFTILANFDSRSRGKXANTLLGSIHTFDEPAACDDSTFQPCSPRALA 300
Db 241 APEILCYLQSFMTGFTILANFDSRSRGKXANTLLGSIHTFDEPAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAAEQLYDALYQMD 360
Db 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAAEQLYDALYQMD 360
QY 361 KQGSLEVTDSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTFADGFSIVETHAASNG 420
Db 361 KQGSLEVTDSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTFADGFSIVETHAASNG 420
QY 421 SMSEQYDKSDGEOLSRDLTWSYAALLTANRRNSVVPASWGETSASSVPGTCAATSAIG 480
Db 421 SMSEQYDKSDGEOLSRDLTWSYAALLTANRRNSVVPASWGETSASSVPGTCAATSAIG 480
QY 481 TVSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTS 529
Db 481 TVSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTS 529

RESULT 9
US-10-418-836-13
; Sequence 13, Application US/10418836
; Publication No. US20040018573A1

GENERAL INFORMATION:
APPLICANT: Power, Scott D.
APPLICANT: Wang, Huaming
APPLICANT: Ward, Michael
TITLE OF INVENTION: Production of Functional Antibodies in
TITLE OF INVENTION: Filamentous Fungi
FILE REFERENCE: GC741-2
CURRENT APPLICATION NUMBER: US/10/418,836
CURRENT FILING DATE: 2003-04-17
PRIOR APPLICATION NUMBER: US 60/373,889
PRIOR FILING DATE: 2002-04-18
PRIOR APPLICATION NUMBER: US 60/411,540
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: US 60/452,134
PRIOR FILING DATE: 2003-03-04
PRIOR APPLICATION NUMBER: US 60/411,537
PRIOR FILING DATE: 2002-09-18
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 763
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion protein
US-10-418-836-13

Query Match 98.7%; Score 2735; DB 15; Length 763;
Best Local Similarity 99.6%; Pred. No. 7.5e-234;
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVAATATLNNIGAGAWVSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVAATATLNNIGAGAWVSGADSGI 60
QY 61 VVASPTDNDYFYTTWRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
DB 61 VVASPTDNDYFYTTWRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
QY 121 DLSGAGLGEPRKNNVDETAATYTGSGWRPQRDGPALRATAMIGFGWLLDNGYSTATDIW 180
DB 121 DLSGAGLGEPRKNNVDETAATYTGSGWRPQRDGPALRATAMIGFGWLLDNGYSTATDIW 180
QY 121 DLSGAGLGEPRKNNVDETAATYTGSGWRPQRDGPALRATAMIGFGWLLDNGYSTATDIW 180
DB 121 DLSGAGLGEPRKNNVDETAATYTGSGWRPQRDGPALRATAMIGFGWLLDNGYSTATDIW 180
QY 181 PLVRNDLSYVAQYQWNOTGYDLMEEVNCSFFTTAVQHRALVBSGAPATAVSSCWCDSQ 240
DB 181 PLVRNDLSYVAQYQWNOTGYDLMEEVNCSFFTTAVQHRALVBSGAPATAVSSCWCDSQ 240
QY 241 APRILCYLGSFMTGSFTLANFSDSSRGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
DB 241 APRILCYLGSFMTGSFTLANFSDSSRGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAEQLYDALYQMD 360
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAEQLYDALYQMD 360
QY 361 KQSLLEVTVDSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
DB 361 KQSLLEVTVDSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
QY 421 SMSEQYDKSDGEOBSARDLTWYSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOBSARDLTWYSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TVSSSVTVSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTYKR 529
DB 481 TVSSSVTVSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTYKR 529

RESULT 10
US-10-418-836-10
; Sequence 16, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
APPLICANT: Power, Scott D.

APPLICANT: Power, Scott D.
APPLICANT: Wang, Huaming
APPLICANT: Ward, Michael
TITLE OF INVENTION: Production of Functional Antibodies in
TITLE OF INVENTION: Filamentous Fungi
FILE REFERENCE: GC741-2
CURRENT APPLICATION NUMBER: US/10/418,836
CURRENT FILING DATE: 2003-04-17
PRIOR APPLICATION NUMBER: US 60/373,889
PRIOR FILING DATE: 2002-04-18
PRIOR APPLICATION NUMBER: US 60/411,540
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: US 60/452,134
PRIOR FILING DATE: 2003-03-04
PRIOR APPLICATION NUMBER: US 60/411,537
PRIOR FILING DATE: 2002-09-18
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10
LENGTH: 979
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion protein
US-10-418-836-10

Query Match 98.7%; Score 2735; DB 15; Length 979;
Best Local Similarity 99.6%; Pred. No. 1.1e-233;
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVAATATLNNIGAGAWVSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVAATATLNNIGAGAWVSGADSGI 60
QY 61 VVASPTDNDYFYTTWRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
DB 61 VVASPTDNDYFYTTWRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
QY 121 DLSGAGLGEPRKNNVDETAATYTGSGWRPQRDGPALRATAMIGFGWLLDNGYSTATDIW 180
DB 121 DLSGAGLGEPRKNNVDETAATYTGSGWRPQRDGPALRATAMIGFGWLLDNGYSTATDIW 180
QY 121 DLSGAGLGEPRKNNVDETAATYTGSGWRPQRDGPALRATAMIGFGWLLDNGYSTATDIW 180
DB 121 DLSGAGLGEPRKNNVDETAATYTGSGWRPQRDGPALRATAMIGFGWLLDNGYSTATDIW 180
QY 181 PLVRNDLSYVAQYQWNOTGYDLMEEVNCSFFTTAVQHRALVBSGAPATAVSSCWCDSQ 240
DB 181 PLVRNDLSYVAQYQWNOTGYDLMEEVNCSFFTTAVQHRALVBSGAPATAVSSCWCDSQ 240
QY 241 APRILCYLGSFMTGSFTLANFSDSSRGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
DB 241 APRILCYLGSFMTGSFTLANFSDSSRGKDANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAEQLYDALYQMD 360
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAEQLYDALYQMD 360
QY 361 KQSLLEVTVDSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
DB 361 KQSLLEVTVDSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
QY 421 SMSEQYDKSDGEOBSARDLTWYSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOBSARDLTWYSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TVSSSVTVSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTYKR 529
DB 481 TVSSSVTVSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTYKR 529

RESULT 11
US-10-418-836-16
; Sequence 16, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
APPLICANT: Power, Scott D.

APPLICANT: Wang, Huaming
APPLICANT: Ward, Michael
TITLE OF INVENTION: Production of Functional Antibodies in
FILE REFERENCE: GC741-2
CURRENT APPLICATION NUMBER: US/10/418, 836
CURRENT FILING DATE: 2003-04-17
PRIOR APPLICATION NUMBER: US 60/373, 889
PRIOR FILING DATE: 2002-04-18
PRIOR APPLICATION NUMBER: US 60/411, 540
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: US 60/452, 134
PRIOR FILING DATE: 2003-03-04
PRIOR APPLICATION NUMBER: US 60/411, 537
PRIOR FILING DATE: 2002-09-18
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 979
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion protein
US-10-418-836-16

Query Match 98.7%; Score 2735; DB 15; Length 979;
Best Local Similarity 99.6%; Pred. No. 1,1e-233;
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
QY 61 VVASPSNDNDYFYTWTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120
DB 61 VVASPSNDNDYFYTWTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120
QY 121 DLSSGAGLGEPEKKNVDETAYTSGMGRPORDPALRATAMIGFGOMLLDNGYSTATDIY 180
DB 121 DLSSGAGLGEPEKKNVDETAYTSGMGRPORDPALRATAMIGFGOMLLDNGYSTATDIY 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTTIAVQHRALVEGSAFAVAGSSCWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTTIAVQHRALVEGSAFAVAGSSCWCDSQ 240
QY 241 APEILCYLQSFMTGSPFLANFDSRSRSGKANTLGSIHFPDPEAACDDSTFQPCSPRALA 300
DB 241 APEILCYLQSFMTGSPFLANFDSRSRSGKANTLGSIHFPDPEAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVNAGRYPEPTYNGNMPFLCTTAAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVNAGRYPEPTYNGNMPFLCTTAAAEQLYDALYQMD 360
QY 361 KQSLLEVTVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSI VETHAASNG 420
DB 361 KQSLLEVTVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSI VETHAASNG 420
QY 421 SMEEYDKSDGEBLSARDLWTSYALLTANNRNSVVPASWGTSSASVPGTCAATSAIG 480
DB 421 SMEEYDKSDGEBLSARDLWTSYALLTANNRNSVVPASWGTSSASVPGTCAATSAIG 480
QY 481 TYSSTVTVSWPSPVATGCTTTTATPTGSGSVTSTKTTATASKTSTYKR 529
DB 481 TYSSTVTVSWPSPVATGCTTTTATPTGSGSVTSTKTTATASKTSTYKR 529

RESULT 12
US-10-418-836-19
; Sequence 19; Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huaming

APPLICANT: Ward, Michael
APPLICANT: Ward, Michael
TITLE OF INVENTION: Production of Functional Antibodies in
FILE REFERENCE: GC741-2
CURRENT APPLICATION NUMBER: US/10/418, 836
CURRENT FILING DATE: 2003-04-17
PRIOR APPLICATION NUMBER: US 60/373, 889
PRIOR FILING DATE: 2002-04-18
PRIOR APPLICATION NUMBER: US 60/411, 540
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: US 60/452, 134
PRIOR FILING DATE: 2003-03-04
PRIOR APPLICATION NUMBER: US 60/411, 537
PRIOR FILING DATE: 2002-09-18
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 738
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion protein
US-10-418-836-19

Query Match 98.2%; Score 2720.5; DB 15; Length 738;
Best Local Similarity 98.5%; Pred. No. 1,4e-232;
Matches 527; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
QY 61 VVASPSNDNDYFYTWTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120
DB 61 VVASPSNDNDYFYTWTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120
QY 121 DLSSGAGLGEPEKKNVDETAYTSGMGRPORDPALRATAMIGFGOMLLDNGYSTATDIY 180
DB 121 DLSSGAGLGEPEKKNVDETAYTSGMGRPORDPALRATAMIGFGOMLLDNGYSTATDIY 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTTIAVQHRALVEGSAFAVAGSSCWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTTIAVQHRALVEGSAFAVAGSSCWCDSQ 240
QY 241 APEILCYLQSFMTGSPFLANFDSRSRSGKANTLGSIHFPDPEAACDDSTFQPCSPRALA 300
DB 241 APEILCYLQSFMTGSPFLANFDSRSRSGKANTLGSIHFPDPEAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVNAGRYPEPTYNGNMPFLCTTAAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVNAGRYPEPTYNGNMPFLCTTAAAEQLYDALYQMD 360
QY 361 KQSLLEVTVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSI VETHAASNG 420
DB 361 KQSLLEVTVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSI VETHAASNG 420
QY 421 SMEEYDKSDGEBLSARDLWTSYALLTANNRNSVVPASWGTSSASVPGTCAATSAIG 480
DB 421 SMEEYDKSDGEBLSARDLWTSYALLTANNRNSVVPASWGTSSASVPGTCAATSAIG 480
QY 481 TYSSTVTVSWPSPVATGCTTTTATPTGSGSVTSTKTTATASKTSTYKR 534
DB 481 TYSSTVTVSWPSPVATGCTTTTATPTGSGSVTSTKTTATASKTSTYKR 534

RESULT 13
US-10-418-836-30
; Sequence 30; Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huaming
; APPLICANT: Ward, Michael

```
; TITLE OF INVENTION: Production of Functional Antibodies in
; FILE OF INVENTION: Filamentous Fungi
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 741
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-30
```

```
Query Match      98.0%; Score 2716; DB 15; Length 741;
Best Local Similarity 98.0%; Pred. No. 3,5e-232;
Matches 527; Conservative 1; Mismatches 6; Indels 4; Gaps 1;
```

```
QY      1 MSFRLSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
Db      1 MSFRLSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
QY      61 VVASPTDNPDYFYTWTRDSGLVLTIVLDFRNGDTSLSLTINNYISAQAIIVGGINPSG 120
Db      61 VVASPTDNPDYFYTWTRDSGLVLTIVLDFRNGDTSLSLTINNYISAQAIIVGGINPSG 120
QY      121 DLSGAGLGEPKENVDETAATYGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSATDIW 180
Db      121 DLSGAGLGEPKENVDETAATYGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSATDIW 180
QY      181 PLVNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVSSCSMCDSQ 240
Db      181 PLVNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVSSCSMCDSQ 240
QY      241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
Db      241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
QY      301 NHRKVVDSPRSITVTLNDGLSDSEAVAGRYPEDTYNGNPFVLCITLAAEQLYDALYQMD 360
Db      301 NHRKVVDSPRSITVTLNDGLSDSEAVAGRYPEDTYNGNPFVLCITLAAEQLYDALYQMD 360
QY      361 KQGSLEVTDVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTADGFSVIVETHAASNG 420
Db      361 KQGSLEVTDVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTADGFSVIVETHAASNG 420
QY      421 SMSBOYDKSDGEQLSARDLTWSTYAAALLTANNRNSVVPASWGTSASVPGTCAATSAIG 480
Db      421 SMSBOYDKSDGEQLSARDLTWSTYAAALLTANNRNSVVPASWGTSASVPGTCAATSAIG 480
QY      481 TYSSSVTWSMPSIVAGCGTTTATPTGSGSVTSTSKTTATASKTS-----TTTRSGMSL 534
Db      481 TYSSSVTWSMPSIVAGCGTTTATPTGSGSVTSTSKTTATASKTS-----TTTRSGMSL 534
```

```
RESULT 14
US-10-418-836-38
; Sequence 38, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huaming
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
```

```
; TITLE OF INVENTION: Filamentous Fungi
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 972
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-38
```

```
Query Match      98.0%; Score 2715.5; DB 15; Length 972;
Best Local Similarity 95.5%; Pred. No. 5.8e-232;
Matches 530; Conservative 0; Mismatches 4; Indels 21; Gaps 2;
```

```
QY      1 MSFRLSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
Db      1 MSFRLSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
QY      61 VVASPTDNPDYFYTWTRDSGLVLTIVLDFRNGDTSLSLTINNYISAQAIIVGGINPSG 120
Db      61 VVASPTDNPDYFYTWTRDSGLVLTIVLDFRNGDTSLSLTINNYISAQAIIVGGINPSG 120
QY      121 DLSGAGLGEPKENVDETAATYGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSATDIW 180
Db      121 DLSGAGLGEPKENVDETAATYGSWGRPQRDGPALRATAMIGFGQWLLDNGYTSATDIW 180
QY      181 PLVNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVSSCSMCDSQ 240
Db      181 PLVNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVSSCSMCDSQ 240
QY      241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
Db      241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
QY      301 NHRKVVDSPRSITVTLNDGLSDSEAVAGRYPEDTYNGNPFVLCITLAAEQLYDALYQMD 360
Db      301 NHRKVVDSPRSITVTLNDGLSDSEAVAGRYPEDTYNGNPFVLCITLAAEQLYDALYQMD 360
QY      361 KQGSLEVTDVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTADGFSVIVETHAASNG 420
Db      361 KQGSLEVTDVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTADGFSVIVETHAASNG 420
QY      421 SMSBOYDKSDGEQLSARDLTWSTYAAALLTANNRNSVVPASWGTSASVPGTCAATSAIG 480
Db      421 SMSBOYDKSDGEQLSARDLTWSTYAAALLTANNRNSVVPASWGTSASVPGTCAATSAIG 480
QY      481 TYSSSVTWSMPSIVAGCGTTTATPTGSGSVTSTSKTTATAS-----KT 524
Db      481 TYSSSVTWSMPSIVAGCGTTTATPTGSGSVTSTSKTTATAS-----KT 524
QY      525 S-----TTTRSGMSL 534
Db      541 SETLSLCTVSGFSL 555
```

```
RESULT 15
US-10-418-836-26
; Sequence 26, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
```

```

; APPLICANT: Wang, Huaming
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
; TITLE OF INVENTION: Filamentous Fungi
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FaestSeq for Windows Version 4.0
; SEQ ID NO: 26
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
; US-10-418-836-26

```

```

Query Match      98.0%; Score 2715; DB 15; Length 739;
Best Local Similarity 99.8%; Pred. No. 4.3e-232;
Matches 522; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MSFRSLALSLGLVCTGLANVSKRATLDSWLSNEATVARTAIINIGADGAWYSGADSGI 60
DB 1 MSFRSLALSLGLVCTGLANVSKRATLDSWLSNEATVARTAIINIGADGAWYSGADSGI 60
QY 61 VVASPTDNDPDYFYWTTRSGVLKTLVDFRNGDTSLSTIENYISAQAIYVGISNPSG 120
DB 61 VVASPTDNDPDYFYWTTRSGVLKTLVDFRNGDTSLSTIENYISAQAIYVGISNPSG 120
QY 121 DLSSGAGLEPKFNVDVETAYTSGMGRPORDGPRALRATAMIGFQWMLDNGYSTATDIYW 180
DB 121 DLSSGAGLEPKFNVDVETAYTSGMGRPORDGPRALRATAMIGFQWMLDNGYSTATDIYW 180
QY 121 DLSSGAGLEPKFNVDVETAYTSGMGRPORDGPRALRATAMIGFQWMLDNGYSTATDIYW 180
DB 121 DLSSGAGLEPKFNVDVETAYTSGMGRPORDGPRALRATAMIGFQWMLDNGYSTATDIYW 180
QY 181 PLVRNDLSYVAQYMNQGYDLMEEVNGSSFFITAVQHRALVEGSAFATAVGSWSCWDSQ 240
DB 181 PLVRNDLSYVAQYMNQGYDLMEEVNGSSFFITAVQHRALVEGSAFATAVGSWSCWDSQ 240
QY 181 PLVRNDLSYVAQYMNQGYDLMEEVNGSSFFITAVQHRALVEGSAFATAVGSWSCWDSQ 240
DB 181 PLVRNDLSYVAQYMNQGYDLMEEVNGSSFFITAVQHRALVEGSAFATAVGSWSCWDSQ 240
QY 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSIHFDPEAACDDSTFPQCSPRALA 300
DB 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSIHFDPEAACDDSTFPQCSPRALA 300
QY 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSIHFDPEAACDDSTFPQCSPRALA 300
DB 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSIHFDPEAACDDSTFPQCSPRALA 300
QY 301 NHEKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360
DB 301 NHEKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360
QY 301 NHEKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360
DB 301 NHEKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360
QY 361 KQGSLEVTDVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420
DB 361 KQGSLEVTDVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420
QY 361 KQGSLEVTDVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420
DB 361 KQGSLEVTDVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420
QY 421 SMSEQYDKSDGEOLSARDLTWSYAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOLSARDLTWSYAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TYSSVVTWSPSIVATGCTTTTATPTGSGSVTSTKTTATASK 523
DB 481 TYSSVVTWSPSIVATGCTTTTATPTGSGSVTSTKTTATASK 523

```

```

RESULT 16
US-10-418-836-27
; Sequence 27, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huaming

```

```

; APPLICANT: Ward, Michael
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
; TITLE OF INVENTION: Filamentous Fungi
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FaestSeq for Windows Version 4.0
; SEQ ID NO: 27
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
; US-10-418-836-27

```

```

Query Match      98.0%; Score 2715; DB 15; Length 739;
Best Local Similarity 98.1%; Pred. No. 4.3e-232;
Matches 526; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

```

```

QY 1 MSFRSLALSLGLVCTGLANVSKRATLDSWLSNEATVARTAIINIGADGAWYSGADSGI 60
DB 1 MSFRSLALSLGLVCTGLANVSKRATLDSWLSNEATVARTAIINIGADGAWYSGADSGI 60
QY 61 VVASPTDNDPDYFYWTTRSGVLKTLVDFRNGDTSLSTIENYISAQAIYVGISNPSG 120
DB 61 VVASPTDNDPDYFYWTTRSGVLKTLVDFRNGDTSLSTIENYISAQAIYVGISNPSG 120
QY 121 DLSSGAGLEPKFNVDVETAYTSGMGRPORDGPRALRATAMIGFQWMLDNGYSTATDIYW 180
DB 121 DLSSGAGLEPKFNVDVETAYTSGMGRPORDGPRALRATAMIGFQWMLDNGYSTATDIYW 180
QY 121 DLSSGAGLEPKFNVDVETAYTSGMGRPORDGPRALRATAMIGFQWMLDNGYSTATDIYW 180
DB 121 DLSSGAGLEPKFNVDVETAYTSGMGRPORDGPRALRATAMIGFQWMLDNGYSTATDIYW 180
QY 181 PLVRNDLSYVAQYMNQGYDLMEEVNGSSFFITAVQHRALVEGSAFATAVGSWSCWDSQ 240
DB 181 PLVRNDLSYVAQYMNQGYDLMEEVNGSSFFITAVQHRALVEGSAFATAVGSWSCWDSQ 240
QY 181 PLVRNDLSYVAQYMNQGYDLMEEVNGSSFFITAVQHRALVEGSAFATAVGSWSCWDSQ 240
DB 181 PLVRNDLSYVAQYMNQGYDLMEEVNGSSFFITAVQHRALVEGSAFATAVGSWSCWDSQ 240
QY 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSIHFDPEAACDDSTFPQCSPRALA 300
DB 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSIHFDPEAACDDSTFPQCSPRALA 300
QY 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSIHFDPEAACDDSTFPQCSPRALA 300
DB 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSIHFDPEAACDDSTFPQCSPRALA 300
QY 301 NHEKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360
DB 301 NHEKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360
QY 301 NHEKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360
DB 301 NHEKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360
QY 361 KQGSLEVTDVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420
DB 361 KQGSLEVTDVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420
QY 361 KQGSLEVTDVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420
DB 361 KQGSLEVTDVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420
QY 421 SMSEQYDKSDGEOLSARDLTWSYAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOLSARDLTWSYAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TYSSVVTWSPSIVATGCTTTTATPTGSGSVTSTKTTATASK -KTSYTRSGMSL 534
DB 481 TYSSVVTWSPSIVATGCTTTTATPTGSGSVTSTKTTATASK -KTSYTRSGMSL 534

```

```

RESULT 17
US-10-418-836-28
; Sequence 28, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huaming
; APPLICANT: Ward, Michael

```

```
; TITLE OF INVENTION: Production of Functional Antibodies in
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
; US-10-418-836-28
```

```
Query Match          98.0%; Score 2714.5; DB 15; Length 740;
Best Local Similarity 98.0%; Pred. No. 4,7e-232;
Matches 526; Conservative 2; Mismatches 6; Indels 3; Gaps 1;
```

```
QY 1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
DB 1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
QY 61 VVASPSTNDPDIYYTWTTRDSGLVLTIVDLFRNGDTSLSLTENYISAQAIIVGGINPSG 120
DB 61 VVASPSTNDPDIYYTWTTRDSGLVLTIVDLFRNGDTSLSLTENYISAQAIIVGGINPSG 120
QY 121 DLSGAGLGEKPKNVDETAATGSGWRPQRDGPALRATAMIGFGQWLLDNGYTSATDIYW 180
DB 121 DLSGAGLGEKPKNVDETAATGSGWRPQRDGPALRATAMIGFGQWLLDNGYTSATDIYW 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVAGSSCWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVAGSSCWCDSQ 240
QY 241 APEILCYLQSFMTGSIILANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
DB 241 APEILCYLQSFMTGSIILANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
QY 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLLAAAEQLDALYQMD 360
DB 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLLAAAEQLDALYQMD 360
QY 361 KQGSLEVTDVSLDFPKALYSDAATGTYSSSSYSSIVDAVKTFADGFVSIIVETHAASNG 420
DB 361 KQGSLEVTDVSLDFPKALYSDAATGTYSSSSYSSIVDAVKTFADGFVSIIVETHAASNG 420
QY 421 SMSEQYDKSDGEQLSARDLTWSYAALLITANNRRNSVVPASWGTSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEQLSARDLTWSYAALLITANNRRNSVVPASWGTSASVPGTCAATSAIG 480
QY 481 TYSSVVTWSPSIYVAGGTTTATPTGSGSVTSKTTATAS--KSTTTTSGMSL 534
DB 481 TYSSVVTWSPSIYVAGGTTTATPTGSGSVTSKTTATASISKRDIDMTQSPSSL 537
```

```
RESULT 18
US-10-418-836-34
; Sequence 34, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huang
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
```

```
; TITLE OF INVENTION: Filamentous Fungi
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
; US-10-418-836-34
```

```
Query Match          98.0%; Score 2714.5; DB 15; Length 740;
Best Local Similarity 98.0%; Pred. No. 4,7e-232;
Matches 526; Conservative 2; Mismatches 6; Indels 3; Gaps 1;
```

```
QY 1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
DB 1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
QY 61 VVASPSTNDPDIYYTWTTRDSGLVLTIVDLFRNGDTSLSLTENYISAQAIIVGGINPSG 120
DB 61 VVASPSTNDPDIYYTWTTRDSGLVLTIVDLFRNGDTSLSLTENYISAQAIIVGGINPSG 120
QY 121 DLSGAGLGEKPKNVDETAATGSGWRPQRDGPALRATAMIGFGQWLLDNGYTSATDIYW 180
DB 121 DLSGAGLGEKPKNVDETAATGSGWRPQRDGPALRATAMIGFGQWLLDNGYTSATDIYW 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVAGSSCWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVAGSSCWCDSQ 240
QY 241 APEILCYLQSFMTGSIILANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
DB 241 APEILCYLQSFMTGSIILANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
QY 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLLAAAEQLDALYQMD 360
DB 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLLAAAEQLDALYQMD 360
QY 361 KQGSLEVTDVSLDFPKALYSDAATGTYSSSSYSSIVDAVKTFADGFVSIIVETHAASNG 420
DB 361 KQGSLEVTDVSLDFPKALYSDAATGTYSSSSYSSIVDAVKTFADGFVSIIVETHAASNG 420
QY 421 SMSEQYDKSDGEQLSARDLTWSYAALLITANNRRNSVVPASWGTSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEQLSARDLTWSYAALLITANNRRNSVVPASWGTSASVPGTCAATSAIG 480
QY 481 TYSSVVTWSPSIYVAGGTTTATPTGSGSVTSKTTATAS--KSTTTTSGMSL 534
DB 481 TYSSVVTWSPSIYVAGGTTTATPTGSGSVTSKTTATASISKRDIDMTQSPSSL 537
```

```
RESULT 19
US-10-418-836-29
; Sequence 29, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huang
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
```


FILE REFERENCE: GC741-2
CURRENT APPLICATION NUMBER: US/10/418,836
CURRENT FILING DATE: 2003-04-17
PRIOR APPLICATION NUMBER: US 60/373,889
PRIOR FILING DATE: 2002-04-18
PRIOR APPLICATION NUMBER: US 60/411,540
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: US 60/452,134
PRIOR FILING DATE: 2003-03-04
PRIOR APPLICATION NUMBER: US 60/411,537
PRIOR FILING DATE: 2002-09-18
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 742
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion protein
US-10-418-836-29

Query Match 98.0%; Score 2714.5; DB 15; Length 742;
Best Local Similarity 97.6%; Pred. No. 4,86-232;
Matches 526; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

QY 1 MSFRLSLALSGVCTGLANVSKRATLDSMLSEATVARTAILNNIGADGAWSGADSGI 60
DB 1 MSFRLSLALSGVCTGLANVSKRATLDSMLSEATVARTAILNNIGADGAWSGADSGI 60
QY 61 VVASPSTNDPDYFYTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAVOGISNPSG 120
DB 61 VVASPSTNDPDYFYTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAVOGISNPSG 120
QY 121 DISSGAGLGEPRKENVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIYW 180
DB 121 DISSGAGLGEPRKENVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIYW 180
QY 181 PLVRNRLSYVAQWNOGTGDLMEEVNGSSFETIAVQHRALVEGSAPATVAGSSCSWCDQ 240
DB 181 PLVRNRLSYVAQWNOGTGDLMEEVNGSSFETIAVQHRALVEGSAPATVAGSSCSWCDQ 240
QY 241 APEILCYLQSFMTGSFILANFDSRSRSGKDNATLGSIHTEPDAACDDSTFOPCSPRALA 300
DB 241 APEILCYLQSFMTGSFILANFDSRSRSGKDNATLGSIHTEPDAACDDSTFOPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABEQLYDALYQMD 360
QY 361 KQGSLEVTQVSLDFFKALYSDAATGYSSSSSTYSSTYSDAVKTFADGFVSIYETHAASNG 420
DB 361 KQGSLEVTQVSLDFFKALYSDAATGYSSSSSTYSSTYSDAVKTFADGFVSIYETHAASNG 420
QY 421 SMSEQYDKSDGRLSARDLTWSYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGRLSARDLTWSYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TYSSVTVTSMPSIVATGTTTATPTGSGSVTSTSKTTATAS 534
DB 481 TYSSVTVTSMPSIVATGTTTATPTGSGSVTSTSKTTATAS 534

RESULT 20
US-10-418-836-31
Sequence 31, Application US/10418836
Publication No. US20040018573A1
GENERAL INFORMATION:
APPLICANT: Power, Scott D.
APPLICANT: Wang, Huang
APPLICANT: Ward, Michael
TITLE OF INVENTION: Production of Functional Antibodies in
FILAMENTOUS FUNGI
FILE REFERENCE: GC741-2

CURRENT APPLICATION NUMBER: US/10/418,836
CURRENT FILING DATE: 2003-04-17
PRIOR APPLICATION NUMBER: US 60/373,889
PRIOR FILING DATE: 2002-04-18
PRIOR APPLICATION NUMBER: US 60/411,540
PRIOR FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: US 60/452,134
PRIOR FILING DATE: 2003-03-04
PRIOR APPLICATION NUMBER: US 60/411,537
PRIOR FILING DATE: 2002-09-18
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31
LENGTH: 743
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion protein
US-10-418-836-31

Query Match 97.9%; Score 2714; DB 15; Length 743;
Best Local Similarity 100.0%; Pred. No. 5,36-232;
Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRLSLALSGVCTGLANVSKRATLDSMLSEATVARTAILNNIGADGAWSGADSGI 60
DB 1 MSFRLSLALSGVCTGLANVSKRATLDSMLSEATVARTAILNNIGADGAWSGADSGI 60
QY 61 VVASPSTNDPDYFYTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAVOGISNPSG 120
DB 61 VVASPSTNDPDYFYTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAVOGISNPSG 120
QY 121 DISSGAGLGEPRKENVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIYW 180
DB 121 DISSGAGLGEPRKENVDETAATYTGSGWRPQRDGPALRATAMIGFQWLLDNGYSTATDIYW 180
QY 181 PLVRNRLSYVAQWNOGTGDLMEEVNGSSFETIAVQHRALVEGSAPATVAGSSCSWCDQ 240
DB 181 PLVRNRLSYVAQWNOGTGDLMEEVNGSSFETIAVQHRALVEGSAPATVAGSSCSWCDQ 240
QY 241 APEILCYLQSFMTGSFILANFDSRSRSGKDNATLGSIHTEPDAACDDSTFOPCSPRALA 300
DB 241 APEILCYLQSFMTGSFILANFDSRSRSGKDNATLGSIHTEPDAACDDSTFOPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABEQLYDALYQMD 360
QY 361 KQGSLEVTQVSLDFFKALYSDAATGYSSSSSTYSSTYSDAVKTFADGFVSIYETHAASNG 420
DB 361 KQGSLEVTQVSLDFFKALYSDAATGYSSSSSTYSSTYSDAVKTFADGFVSIYETHAASNG 420
QY 421 SMSEQYDKSDGRLSARDLTWSYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGRLSARDLTWSYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TYSSVTVTSMPSIVATGTTTATPTGSGSVTSTSKTTATAS 522
DB 481 TYSSVTVTSMPSIVATGTTTATPTGSGSVTSTSKTTATAS 522

RESULT 21
US-10-418-836-35
Sequence 35, Application US/10418836
Publication No. US20040018573A1
GENERAL INFORMATION:
APPLICANT: Power, Scott D.
APPLICANT: Wang, Huang
APPLICANT: Ward, Michael
TITLE OF INVENTION: Production of Functional Antibodies in
FILAMENTOUS FUNGI
CURRENT APPLICATION NUMBER: US/10/418,836

```
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-35
```

```
Query Match          97.9%; Score 2714; DB 15; Length 743;
Best Local Similarity 100.0%; Pred. No. 5,3e-232;
Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWSGADSGI 60
D 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWSGADSGI 60
QY 61 VVASPTDNDPYFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
D 61 VVASPTDNDPYFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
QY 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFQWLLDNGYSTATDIW 180
D 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFQWLLDNGYSTATDIW 180
QY 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFQWLLDNGYSTATDIW 180
D 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFQWLLDNGYSTATDIW 180
QY 181 PLVNDLSYVAQVWNOGTGYDLMEEVNCGSPFTIAVOHRALVEGSAFATAVGSSCSWCDQ 240
D 181 PLVNDLSYVAQVWNOGTGYDLMEEVNCGSPFTIAVOHRALVEGSAFATAVGSSCSWCDQ 240
QY 241 APEILCYLQSFMTGSFLIANFDSRSGKDNATLLGSIHTFPPEACDSTFQPCSPRALA 300
D 241 APEILCYLQSFMTGSFLIANFDSRSGKDNATLLGSIHTFPPEACDSTFQPCSPRALA 300
QY 301 NHKEVDSFRSITLNDGLSDSEAVAVGRYPEDTYNGNPFCTLAQAQLYDALYQMD 360
D 301 NHKEVDSFRSITLNDGLSDSEAVAVGRYPEDTYNGNPFCTLAQAQLYDALYQMD 360
QY 361 KQGSLEVTDVSLDPFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSI VETHAASNG 420
D 361 KQGSLEVTDVSLDPFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSI VETHAASNG 420
QY 421 SMSEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
D 421 SMSEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TYSSVVTYSWPSIVATGCTTTTATPTGSGSVTSTSKTTATAS 522
D 481 TYSSVVTYSWPSIVATGCTTTTATPTGSGSVTSTSKTTATAS 522
```

```
RESULT 22
US-10-418-836-39
; Sequence 39, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huaming
; TITLE OF INVENTION: Production of Functional Antibodies in
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; CURRENT FILING DATE: 2003-04-17
```

```
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-39
```

```
Query Match          97.9%; Score 2714; DB 15; Length 975;
Best Local Similarity 100.0%; Pred. No. 8e-232;
Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWSGADSGI 60
D 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWSGADSGI 60
QY 61 VVASPTDNDPYFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
D 61 VVASPTDNDPYFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120
QY 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFQWLLDNGYSTATDIW 180
D 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFQWLLDNGYSTATDIW 180
QY 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFQWLLDNGYSTATDIW 180
D 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFQWLLDNGYSTATDIW 180
QY 181 PLVNDLSYVAQVWNOGTGYDLMEEVNCGSPFTIAVOHRALVEGSAFATAVGSSCSWCDQ 240
D 181 PLVNDLSYVAQVWNOGTGYDLMEEVNCGSPFTIAVOHRALVEGSAFATAVGSSCSWCDQ 240
QY 241 APEILCYLQSFMTGSFLIANFDSRSGKDNATLLGSIHTFPPEACDSTFQPCSPRALA 300
D 241 APEILCYLQSFMTGSFLIANFDSRSGKDNATLLGSIHTFPPEACDSTFQPCSPRALA 300
QY 301 NHKEVDSFRSITLNDGLSDSEAVAVGRYPEDTYNGNPFCTLAQAQLYDALYQMD 360
D 301 NHKEVDSFRSITLNDGLSDSEAVAVGRYPEDTYNGNPFCTLAQAQLYDALYQMD 360
QY 361 KQGSLEVTDVSLDPFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSI VETHAASNG 420
D 361 KQGSLEVTDVSLDPFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSI VETHAASNG 420
QY 421 SMSEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
D 421 SMSEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TYSSVVTYSWPSIVATGCTTTTATPTGSGSVTSTSKTTATAS 522
D 481 TYSSVVTYSWPSIVATGCTTTTATPTGSGSVTSTSKTTATAS 522
```

```
RESULT 23
US-10-228-063-45
; Sequence 45, Application US/10228063
; Publication No. US20030135885A1
; GENERAL INFORMATION:
; APPLICANT: Lananan, Mike
; TITLE OF INVENTION: Self-processing Plants and Plant Parts
; FILE REFERENCE: 109846.317
; CURRENT APPLICATION NUMBER: US/10/228,063
; CURRENT FILING DATE: 2002-12-12
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 1095
```


Db 13 AFOAALGELDPLEKHSIDIKR-SVDSYIQTETPIAOKNLCNIGASGCRASGAAGVV 71
QY VASPSPTNDPDY-----FYWTROSGVLKTLVDFRNG-DTSLI 99
Db 72 VASPSASSDPLXVEIAQCVSLMAALVTSIDWYTRDAALVTKLIIVDEFTINDVYTTLO 131
QY 100 STIENYISAQAIWQGISNPSGDLSSGAGLGEPEKFNVDETAYTGSWGRPORHGPALRATAM 159
Db 132 NTIQAVAAAOAKLQGVSNPSGSLSNAGAGLGEPEKFNVDLQOFTGAKWRPQRDGPPLALIAL 191
QY 160 IGFQGLLNDNGYITATDITWPLVRNDLSVYAQYKQGYDLMEEVNGSSFFFTIAVOHRA 219
Db 192 IGGKMLVSNQYADTFKSIIMPIVKNDLAVTAAQYNNMTGFDLMEEVNGSSFFFTIAASHRA 251
QY 220 IVEGSAFATAVSGSSGCDQAPELICYOSFWTGS-FILANPDSRSGDANTLIGSH 278
Db 252 LVEGSAFATAVSGSSGACDIAPOILCFQOSFWNSGYYIISNVNRSGDINSVLTSH 311
QY 279 TEDPEAACDSTFQPCSPALANHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNG 338
Db 312 NNDPAGCQVNTFQPCSDRALANHKVVDMSR-FMGVNSRTAGKAAVGRVADVYNG 370
QY 339 NPEWFLCTLAABEOLYALYQMDKQSLLEVTVSLDFPKALYSDAATGTVSSSSSTYSIV 398
Db 371 NPEWFLCTLAABEOLYALYQMDKQSLLEVTVSLDFPKALYSDAATGTVSSSSSTYSIV 430
QY 399 DAVKTFADGFSIVERTHAASNGSMFOYKSDDEQLSARLTVSYAALLTANRRNSVVP 458
Db 431 NNAVYADGFDVDAVYTPSDGSLAOPDDSGAPLSATHLTWSYASFSAARRAGIYV 490
QY 459 ASWGETSASSVPQTCATSAIGTVSSVTVSWPSIVATGTTT-ATPTGSGS-----VT 512
Db 491 PEMGASANSLSPPSCASATVAGSYATATNTSPANLTPASTVTPPTQICGADHDEVLT 550
QY 513 STSKTATASKT 524
Db 551 FNEKVTTSYQOT 562

RESULT 28
US-10-281-673-3
; Sequence 3, Application US/10281673
; Publication No. US20030148452A1
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Key
; APPLICANT: Kimberly Brown
; TITLE OF INVENTION: Promoters For Expressing Genes In A
; FILE REFERENCE: 5611.200-US
; CURRENT APPLICATION NUMBER: US/10/281,673
; PRIOR FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US/09/534,407
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/274,449
; PRIOR FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Fusarium
; US-10-281-673-3

Query Match 51.4%; Score 1425.5; DB 14; Length 581;
Best Local Similarity 52.2%; Pred. No. 1.3e-117;
Matches 273; Conservative 81; Mismatches 146; Indels 23; Gaps 6;

QY 9 LSGVLVCTGL--ANVI---SKRATLWSLSEATVAARAIANNIGADGAWVSGADSGIVVA 63
Db 6 LVLGVALWQGVVSPSKDNLERITQADISIKIGVIANIGAGKRAQGAFAFVVA 65
QY 64 SPSTNDPDYFYWTROSGVLKTLVDFRNGDTSLSTIENYISAQAIWQGISNPSGDLSS 123

Db 66 SPSKEDPDYWTWTRDSALTYKLVVERFIHDKSLQRIKIDEVYSAQAKLQGTTPSGSDE 125
QY 124 SGAGLGEPEKFNVDETAYTGSWGRPORHGPALRATAMIGCOWLNDNGYSTANDIYWPV 183
Db 126 SG-GLGEPEKFNVDETAYTGSWGRPORHGPALRATAMIGCOWLNDNGYSTANDIYWPV 184
QY 184 RNDLSVYAQYKQGYDLMEEVNGSSFFFTIAVOHRAVLEGSAPATVAGSSGCDQAPE 243
Db 185 EKDLATYTFKFNRTGVDLMEEVNGSSFFFTLSASHRLVGAALAKLGLGSCPCVTNNAR 244
QY 244 IICYOSFWTGSFILANP--SSRSGDANTLIGSHITFPDEAACDSTFQPCSPALAN 301
Db 245 VICEFTQFTWGTGVDSININWCKRGLDVNSILSIHTFDPNKCTDSTFQPCSPALAN 304
QY 302 HKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPEWFLCTLAABEOLYALYQMDK 361
Db 305 HKAVDSFRSIYGVNRRQGAAGVRYSEVYDGNEMWYATTLAAABOLYAAVYQMDK 364
QY 362 QGSLLEVTVSLDFPKALYSDAATGTVSSSSSTYSIVDAVKTADGFSIVERTHAASNGS 421
Db 365 LGAIVYDDVLSLFFKDIYVERKYSKITYAKTKITYKEIKAKTYADGFVAVVQYTRKDS 424
QY 422 MSEQYKSDGEOLSARDLTWSYAAALLTANRRNSVVPASWGETSASSVPQTCATSAIGT 481
Db 425 LAEQFDKSTGAPKSAVHLTWSYAAFAVATTBRRDGIISPWGESSANKVPACQADACDT 484
QY 482 -----YSSVTVTSWPSIVATGTTT-----TPTGSG 509
Db 485 TTFPSVKNQVSSDQKYVYVVGVTLELNSWSPDGLALTPSSSG 527

RESULT 29
US-10-213-990-21
; Sequence 21, Application US/10213990
; Publication No. US20030082595A1
; GENERAL INFORMATION:
; APPLICANT: Bussey, Bo
; APPLICANT: Jiang, Bo
; APPLICANT: Storms, Reg
; APPLICANT: Roemer, Terry
; TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUMIGATUS ENCODING INDUSTRIAL
; FILE REFERENCE: 10182-019-999
; CURRENT APPLICATION NUMBER: US/10/213,990
; PRIOR FILING DATE: 2002-08-05
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 704
; TYPE: PRT
; ORGANISM: Aspergillus
; US-10-213-990-21

Query Match 50.8%; Score 1408.5; DB 14; Length 704;
Best Local Similarity 52.0%; Pred. No. 5.7e-116;
Matches 268; Conservative 85; Mismatches 139; Indels 23; Gaps 5;

QY 27 LPSWLSNEATVARTALANNIGADGAWVSGADSGIVASPSPTNDPDYFYWTROSGVLKLT 86
Db 118 LVSWLAQETSYALDGLVNLNVGPNKATGASSGIIIASPSQSNPDYFYWTRODALTVKY 177
QY 87 LVDLF-RNGDTSLSTIENYISAQAIWQGISNPSGDLSSGAGLGEPEKFNVDETAYTGSW 145
Db 178 LVGSFADHDPAIQRITIEDVVSQALQTVSNPSGMLSSG-GLGEPEKLRVDSAFHGSWG 236
QY 146 RPDORPALRATYAMIGFQWLDNGYTSATDITWPLVRNDLSVYAQYKQGYDLMEEV 205
Db 237 RPSDDBPALRATITLISYALMIDNGYTSVTSWPIYIIONDSLYLEFNNSSFFDLMEEV 286
QY 206 NGSSFFFTIAVOHRAVLEGSAPATVAGSSGCDQAPELICYOSFWTGSFILANPDSR 265
Db 297 RGSSFFFTIAVOHRAVLEGSAPATVAGSSGCDQAPELICYOSFWTGSFILANPDSR 356

DB 424 HOLIDAYRTYADGFMKRV 441

RESULT 32

US-10-369-493-12533
Sequence 12533, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 12533
LENGTH: 432
TYPE: PRT
ORGANISM: Aspergillus nidulans
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(432)
OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-12533

Query Match Best Local Similarity 40.9%; Score 1132; DB 15; Length 432;

Matches 210; Conservative 50; Mismatches 73; Indels 2; Gaps 1;

QY 200 DLWEVNGSSFFITIAVQHRALVEGSAFATVAVGSSGSCWDSQDAPEIICYLQSFMTGSFILA 259
DB 1 DLGXHDMSFFIVVTHSLVVGSTFARTVGASCPKCDQAQIICYMNFMTGSITNA 60
QY 260 NFDSSSGKDANTLLGSIHTFDEACDDSTFQPCSPRALANHKEVDSFRSIYTLNDGL 319
DB 61 NCGGSGSGKDANTLLGSIHTFDEACDDSTFQPCSPRALANHKEVDSFRSIYTLNDGL 120
QY 320 SDEAVAVGPRDPTVYNGPMPFLCTAAEOLYDALYQMDKXGSELTVDVSLDFPKALY 379
DB 121 AQCVAAGAAGYPPDSYNGPMPFLCTAAEOLYDALYQMDKXGSELTVDVSLDFPKALY 180
QY 380 SDAATGTVSSSSSTYSIVDAVKTADGPFYIVETHAASNGSMSEQYDKSDGQSLARDL 439
DB 181 SSAAVGTYASGSAFRAIITAVTYADGVYIVKAAHMANGSISEQPKTYGTCVSARDL 240
QY 440 TWSYALLTANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTVTSWPSIVATGCT 499
DB 241 TWSYALLTANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTVTSWPSIVATGCT 300
QY 500 TTTATPTGSGSVSTSTKTATASKTSTTTSGMSL 534
DB 301 SATTTTMAITSTATSSSTTSTT--TACTTPTVAV 333

RESULT 33

US-10-369-493-2490
Sequence 2490, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 2490
LENGTH: 450
TYPE: PRT
ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2490

Query Match Best Local Similarity 28.2%; Score 781.5; DB 15; Length 450;

Matches 179; Conservative 68; Mismatches 166; Indels 63; Gaps 12;

QY 7 LALSLVCTG--LANVISKRATLDSWLSNEATVARTAILNIGADGAWVSGDSGLVVA 63
DB 8 LILGVVASAESLSPKRSKEASMDWETDOQKIAMGHMLNIGSGMNAKDINPCIIA 67
QY 64 SPSTNDPDTYQWVRDSCGLVLTVDLPFRNGDTSILSTIENTYSAQAIYQGISNPSGDS 123
DB 68 SPSTNDPDTYQWVRDSCGLVLTVDLPFRNGDTSILSTIENTYSAQAIYQGISNPSGDS 127
QY 124 SGAGLGEPRFVNDYATYTSWGRPQRDGPALATAMIGCGMLDNGYSTATDIWPLV 183
DB 128 AG-GLGEPRFVNDYATYTSWGRPQRDGPALATAMIGCGMLDNGYSTATDIWPLV 186
QY 184 RNDLSYVAQYQWGTGYDLMEVNGSSFFITIAVQHRALVEGSAFATVAVGSSGSCWDSQAP 243
DB 187 LADLDYTAHMTAEASFDLWEEIKDHYFTLAVQKRAMODGTAFKRGIG-----APD 237
QY 244 -----ILCYQSFV--TGSFILANPDS--SRSGKDANTLLGSIHT--FDPEACD 287
DB 238 QAAIYQRTPEIDLKGEFMDPEMGVYKGGKRVDSGLDCSTLLSLVSNFDF----- 291
QY 288 DSTFQPCSPRALANHKEVDSFRSIYTLNDGLSDSEAVAVGPRDPTVYNGPMPFLCTAA 342
DB 292 -----WHILPFLKQETMTRODPVAVQGMKQ-----AMGRPELVYDGVSKSIGNPWF 339
QY 343 LCTLAABEOLYDALYQMDKXGSELTVDVSLDFPK--ALYSDAATGTVSSSSSTYSIVDA 400
DB 340 ICTSSAAEIIYKAAIAYYDKNGJPELTBYVINHFEMKFAEFGD-----PYWMSYIRKN 390
QY 401 VTFPAGPFYSIVETHAASNGSMSEQYDKSDGQSLARDL.TWSYALLTANNRNSV 456
DB 391 MRYAADNPLKAAVAFQHPNGSMSEQYDKSDGQSLARDL.TWSYALLTANNRNSV 446

RESULT 34

US-10-228-063-49
Sequence 49, Application US/10228063
Publication No. US20030135885A1
GENERAL INFORMATION:
APPLICANT: Lananan, Mike
TITLE OF INVENTION: Self-processing Plants and Plant Parts
FILE REFERENCE: 109846.317
CURRENT APPLICATION NUMBER: US/10/228,063
CURRENT FILING DATE: 2002-12-12
NUMBER OF SEQ ID NOS: 60
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 49
LENGTH: 579
TYPE: PRT
ORGANISM: Rhizopus oryzae
US-10-228-063-49

Query Match Best Local Similarity 26.2%; Score 725; DB 14; Length 579;

Matches 165; Conservative 71; Mismatches 170; Indels 42; Gaps 10;

QY 25 ATLDSWLSNEATVARTAILNIGADGAWVSGDSGLVVA SPSTNDPDTYQWVRDSCGLVLT 84

Db 143 STISSWIKROEGISRFAMLANINP-----PGSATGFIASAISTAGPPYYTAWRDAALTS 197
Qy 85 KTLVDLFR---NGDTSLLSTIENYISAQAIVOGISNGBLSGAGLGEPRKFNVDETAYT 141
Db 198 NVLIYEVNNTLISGNKTLINLTKDVTFESVKTQSTSTVCN-----CLGEPRKFNVDAGYT 251
Qy 142 GMSGRPRDRPALRATMIGFG-OMLLDNQTSFATITVWPLVRNDLSYVAQVWNOGYD 200
Db 252 GAAGRPNQDGPAAEATYTFILFADSYLQTKDASVYTGTLKPAIFKDLDYVWVWNSNCFD 311
Qy 201 LMEEVNGSSFFITIAVQHRALVEGSAFATAVGSS--CSMCDSOAPILICYOSFMTGSFIL 258
Db 312 LMEEVNGVHFTYTLMMWKGILLADPAKRNGDSIRASTYSTASTANKISSEFVSSNNW 371
Qy 259 ANFDS-----SSSGKADANTL---LGSIHTEPEAACDDSTFOPCSPRALANHEVYDS 308
Db 372 IQVSQSVTVGVSXKGLDVSTLLANLGSV-----DDGFPTGSEKILATAVAVEDS 422
Qy 309 FRSTYTLNDGLSDSEAVAVGRYPEDTY-----YNGNPFICTLAAEQLYDALYQMDKOG 363
Db 423 FASLYPIPNKMLPSYLGNSISIRYPEDTYNGNGNSQNSWFLAVTGAYELYYRAIKEWING 482
Qy 364 SLEVTDSLDFPKALYSDAATG--TYSSSSSTYSIYDAVTFADGPFVSIYETHAASNGSM 422
Db 483 GVTYVSSLSLPEFKKFDSSATSCKYTVGTSDFNMLAQNIALADRFLSTVQLHANHNSGL 542
Qy 423 SEQYDKSDGEOLSNRDLTWSGYAALLTAN 450
Db 543 AEEFRTTGLSTGARDLTWSHSLITAS 570

RESULT 35
US-10-369-493-22096
; Sequence 22096, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22096
; LENGTH: 549
; TYPE: PRT
; ORGANISM: *Saccharomyces cerevisiae*
US-10-369-493-22096

Query Match 21.5%; Score 597; DB 15; Length 549;
Best Local Similarity 30.4%; Pred. No. 5, Se-44;
Matches 153; Conservative 89; Mismatches 202; Indels 60; Gaps 14;

Qy 1 MSFRSLALSLGLVCTGLANVSKRATLDSWLSNEATYARAILINIGADGAWTSGADSGI 60
Db 56 VQURDAVLMNGTVAYD-SNAMDSSALEEWLQCKKXISIKIFENIGPSAVPS-ISPGV 113
Qy 61 VVASPTDNDPFTYTWTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGISNPSG 120
Db 114 VVASPTDNDPFTYTWTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGISNPSG 167
Qy 121 DISSGAG-----LGEPRKFNVDETAYTSGWRPDRDGPALRATMIGFQWILNDG--- 170
Db 168 TLGAGIGYNDTVAGLGPKNVNDTAFTEBWRGRONDGPALRSIALIKIIDIYIKQSGTDL 227
Qy 171 -----YTSATDYLWPLVRNDLSYVAQVWNOGYDLMEEVNGSSFFITIAVQHRALVEGSA 225

Db 228 GAKTTPQSTRA-DIFDDIYKMDLAFITIDHNSSGCFDLMEEVNGHMFETLLVQLSAVDKSL 286
Qy 226 FATAVGSSGSCWD---SOAPILICYL-----OSFMTGSFILANF-----DSSRSCKDA 270
Db 287 YFASRRSSPFVELMQTRDISKFLVDPANGFINKY---NVIQTPMILADTLRSGLDI 343
Qy 271 NTLLGSIHTFDPACADDSTFOPCSPRALANHEVYDSRSIYTLNDGLSDSEAVAVGR 330
Db 344 STLLANVTYHIDASA-SHPFDINDPAVNTLHLHLHMRKSIYPIINDSSKNATGIALGRY 402
Qy 331 PEDTY-----YNGNPFICTLAAEQLYDALYQMDKOGSLEVTDSLDFPKALYSDAATG 385
Db 403 PEDYVYCYGRGSGNPVYLTCTHSTLYQLIYRHISEQHDLYVPANNDGSNAFWSGLVRS 462
Qy 386 TYSS-----SSSTYSIYDAVTFADGPFVSIYETHAASNGSMSEQYDKSDGEG 433
Db 463 NLTTLNDNGSYLLERNTAFPNQTIKIQGLADSLVYLKAAVGTGDELSEGNKXTYTFM 522
Qy 434 LSARDLTWSGYAALLTANNRNSVY 457
Db 523 QGAQHLLTWSYTFWADYQIROEVL 546

RESULT 36
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.0114
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12713
; LENGTH: 2344
; TYPE: PRT
; ORGANISM: *Staphylococcus aureus*
US-09-815-242-12713

Query Match 5.9%; Score 164.5; DB 9; Length 2344;
Best Local Similarity 20.6%; Pred. No. 0.00014;
Matches 116; Conservative 103; Mismatches 234; Indels 111; Gaps 19;

Qy 7 LALSGLVCTGLANVSKRATLDSWLSNEATYARAILINIGADGAWTSG--ADSGIVAS 64
Db 829 LVNSQSVSSMWSGVSXKSLSDPISNSSSTKESBSVSTSDSLTSTLSDSVSMST 888
Qy 65 PSTDNDPFTYTWTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGISNPSGDLSS 124


```

Oy      269  DANTLLGSIHTFPEACDSTFQPCSPRALANHK 303
          |||::|||::|||::|||::|||::|||::|||::|||
Db      1  DAVISLGSIHFTPDAGCGDDSTFQPCSPRALANHK 35

RESULT 38
US-10-172-502-4
; Sequence 4, Application US/10172502
; Publication No. US20030185833A1
; GENERAL INFORMATION...
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2283
; TYPE: PRP
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-4

Query Match      5.8%; Score 159.5; DB 14; Length 2283;
Match Local Similarity 20.2%; Pred. No. 0.00037;
Matches 111; Conservative 106; Mismatches 234; Indels 99; Gaps 16;

17  LANVISIKRATIDSLWLSNE-ATVARTALINNG-----ADGAMVSGADSGIVVASPTDNP 70
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db      815  LBSDEVSAKSKSLSTESNSVSSSTSTSLVNEGVSASSMSDASAKSKTSLSDSISNSSSTKXS 874

```

```

Qy 131 PKFVDETATYIGSMGRFQPDGPALPATAMIGFQWLLDNGTSTATDIWPLVLRNDLSY 190
Db 926 STSNAISTSTS-----LBSASTSDSISINSIASQSAST 961
Qy 191 AQVYNOGYDLMEEVNGSSPFTLIAVQHRALVEGAPATAVGSQSCWCSQAPELLCYLQS 250
Db 962 SKDSQSSTSTSLSTSDSKSMST---SBSLSDSTSTSGSVGSLIASQS----- 100808
Qy 251 FWTSFILANFDSSRSRSGKANTLLIGSIHTFDPPEACDSTFPOCSPRALANHKVEVDSFR 310
Db 1009 -----VSTSTSDSMSTSEIYVSDISTGSSLSASQSKMSVSSMSSTSQSGSTSELS 1066
Qy 311 SIYLANDGLDSEAVAVAGRPEDTYVNGNPMWELCTLAAEOLYALYOMDKOSLEVTDV 370
Db 1061 DSQSTSD--SDSKSL-----QSTSQSGSTSTSTSTASVSTSES---QSTSGMSASQS 1110
Qy 371 SLDFKALYSPAALGCTYSSSSSTSYSTIVDAKTPADGVSIVERTHAASNG----- 420
Db 1111 DSWSISTFSDBSTSDSKASTASESISQASASTSCTSVSTSTSLSTNSERTSTSMDS 1170
Qy 421 -----SMSEQDYDSD--GEOLSGARDLWYSYAALLTANRRNSVVPASGETSASV 469
Db 1171 TSLSTSESDSISTSESTSDSISALIASSEST--FISLSESNSTSDSEBSQASAFLEBSL 1228
Qy 470 PGTCATSAIGTYSVTVTWSPIVATGCTTTA---TPTGSGV---TSTGKTTATASK 523
Db 1229 ESTSESTSESVSSSTSESTSLSDSTSESGSTSLNSGSGSTSTSTSTSISESTPEKS 1288
Qy 524 TSTTRSGMS 533
Db 1289 ESVSTSLMS 1298

```

RESULT 39
US-10-282-122A-43924
; Sequence 43924, Application US/10282122A

```
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,1222A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 43924
LENGTH: 2271
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-10-282-122A-43924

Query Match      5.7%; Score 159; DB 12; Length 2271;
Best Local Similarity 20.4%; Pred. No. 0.00041;
Matches 116; Conservative 105; Mismatches 263; Indels 84; Gaps 17;

QY 21 ISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADGIVASPTDNDPYFYTWTRDS 80
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 748 VTNNMSWDS-VSTSGSTQOGOSVSTSKADDSASTSGSIVVSTASTSKTSVLSLSDS 806
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 81 GLVLKTLVDLFRNGDTLSLTENYISAQAI1VQGISNPSGDSLSSAGLGEKPFNVDETAY 140
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 807 VSASKSLSTSESNVS8--STSTSLVNSQSV---SSSMSCSVSKSTSLSDSISNNGTEK 861
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 141 TGSWGRQGRQOPALRAAMIGFGWMLDNGTSTATTIWPPL-----VRNDLSVAQYWN 195
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 862 SESLSLSTSD--SLRTSTSLSDSLSMSTSGLSKSQSLSTISGSSSTASLSDSTSNAI 919
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 196 QTCGDLMEEVNGSFFTIAYOHRALVEGSAFATAVVGSCWCDGQADEI-----LC 246
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 920 STSLSLSEASLSTSLIS-----NSTANSQASLSTKSQSDQSISLISLSDSKSMS 970
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 247 YLGSF-----WTGSFLA-----NPDSSRSGKDANTLLGSIHTFDEPAACDDSTFQ 292
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 971 TSESLDSTSTSGSVGSLIASQSVSTSDSMSTSEIYSDSISTSGSLSADSKSMS 1030
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 293 PCSPRALANHKEVVDSEFRSIYTLNDGISDSEAVAVGVPEDTYNGNPFWFLCTIAAAEQ 352
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

```
Db 1031 VSSSMSTSGSGSTSELSDSQSTSD--SDSKSL-----STSGSGSTSTSTASAVRT 1083
QY 353 YDALYOMDKXGSLVETDVSLDFPKALYSDAATGYSSSGSTYSSIVDAVTKPDGFVSIY 412
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1084 SES---QSTSGSVASQDSMSISTSFSDSTSDSKASASTASSSISQASASTSGSVSTG 1140
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 413 ETHAASNG-----SMSEQDXSD--GEQLSADLTMSYALLTANN 451
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1141 TSLSTNSERTSTSVSDSTSLSTSESDSISESTSTSDSISEAISABEST--SISLSESN 1198
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 452 RNSVVPASMGERTSASVPTCATSAIGTYSVTVSNPSIYATGCTTTA---TPTGS 508
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1199 TSDSESOASAFISELSESTSESTSESVSSSTSEBSTSLSDSTSESGSTSTSLSNSTGS 1258
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 509 GSV---TSTSKTTATASKTSTTRSGMS 533
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1259 ASISTSTSISESTSTFKSESVSTSLMS 1286
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 40
US-10-369-493-22616
Sequence 22616, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 22616
LENGTH: 1283
TYPE: PRT
ORGANISM: Schizosaccharomyces pombe
US-10-369-493-22616

Query Match      5.7%; Score 158; DB 15; Length 1283;
Best Local Similarity 21.0%; Pred. No. 0.00021;
Matches 118; Conservative 76; Mismatches 244; Indels 124; Gaps 24;

QY 5 SLLALSGLVCTGLANVISKRAITDSWLSNEA--TVARTAILNNIGADGAW--VSGADSGIV- 61
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 180 SILPHSTISTSLSTVSTINDTSASLSKTTSPTAGITTEIVSGSVGYSTFPASGTTGTYE 239
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 62 VASPTDNDPYFYTWTRDSGLVLTVDLFRNGDTLSLTENYISAQAI1VQGISNPSGD 121
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 240 VVEPT-----AGTITETIV---GSSVGYSTFPANGTSGVEVEVEPTAG- 281
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 122 LSSAGLGEKPFNVDETAYTGSWGRPDGDPALRAAMIGFGWMLDNGTSTATTIWP 181
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 282 -----TITETIVSGS-----VGYSTFPANGTSGVEVEPTAG- 313
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 182 LVANDLSYAQYNNQGYDLMEEVNGSFFTIAYO-----RALVEGSAKAT---AV 220
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 314 TACITETIVVS--GSVGYSTFPANGTSGVEVEPTAGTTEITIVSGSVGYSTFPAS 371
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 231 GSSCSMCDGQAPETLCYLOSFWTGS--FIILANPSSRSGKDANTLLGSIHTFDEPAACDDG 289
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 372 GTTSGTVEVEVEPTAGITTEIVSGSKAFSTFPAN-----GTTSGTVEVEVEPTAGIT 425
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 290 TPOPCSPRALANHKEVVDSEFRSIYTLNDGISDSEAV---AVGRYPEDTYNGNPFWFLCTL 346
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 426 TIYSGS-----KFTSTFPANGTSGTVEVEVEPTAGITTE--TIYSGSVGYTSTF 473
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 347 AAAEQLYDALYOMDKXGSLVETDVSLDFPKALYSDAATGYSSSGSTYSSIVDAVTKTFAD 406
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```

Db 474 PA-----NGTSGTVEVEPT-----AGTITETIVSGSKTFTSPASGT-TS 515
Qy 407 GVSIVETHAASGMSSEQ---DKSDGROLSARDLTWMAALL--TANNRNSVVPAM 461
Db 516 GVEVEVEPTA---GTTETIVSGSKAFTSTFPANGTSGTVEVEPTAGTITETIVSGSV 572
Qy 462 GETSASSVPCTCAATSAICTYSSVT-----SMPISVATGCTTTATPCTGSG 509
Db 573 GYTSTFPAGTTSCTVEVEPTAGTITETIVSGSVGTSTFPASGTSGTVEVEPT-AG 631
Qy 510 SVTSTSKTATATASKSTSTTRSG 531
Db 632 TVERI-VGSGVGTSTFPASG 652

RESULT 41
US-10-185-990-10
; Sequence 10, Application US/10185990
; Publication No. US20030073109A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Jae-Gu
; TITLE OF INVENTION: JAE GU PAN ET AL
; FILE REFERENCE: 02589.000100
; CURRENT APPLICATION NUMBER: US/10/185,990
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1621
; TYPE: PRF
; ORGANISM: BACILLUS SUBTILIS
US-10-185-990-10

Query Match 5.6%; Score 155.5; DB 14; Length 1621;
Best Local Similarity 23.6%; Pred. No. 0.00051;
Matches 137; Conservative 66; Mismatches 245; Indels 133; Gaps 27;
Qy 32 SNEATVARTAILNINIGADG-----AMVSGADSGIVVASTDNDP-----YYTWT 77
Db 200 SNETAGNHSDDL IAGYSTGTAGYSTGTQSGEDSSLTAGYSTGTQTAOEGSNLTAGYSTGT 259
Qy 78 --RDGSLVKTIVDLFRNGDTSLSSTIENYISAQAIVOGISNPSGDLSSG-----A 126
Db 260 AGSDSL IAGYSTGTQSGEDSSLT A---GYSTQTQOEGSNLTAGYSTGTACVDSLLA 316
Qy 127 GLGEPEFNDETAYTGSWGRPQRDGPALRATAMIGFQWMLDNGYSTATDIWPLVRND 186
Db 317 GYSTQTQSGDSAL IAGYSTGTQ-----TAQEGSN---LTAGYSTGT-----AGSD 359
Qy 187 LSVVAQYMNQGTGYDLMEVNGSFFITIAV-QHRALVEGSAFATAVGSS-CSMCDQAPET 244
Db 360 SSLLA-----GYGSTQSGDSLSLTAGYSTGTQTAOEGSNLTAGYSTGTAGVDSLLIAG 413
Qy 245 LCVLGSPWTSFPLANFDSRSRSGKDANTL---LGSHTFPBAACDSTFQPC-SPRALA 300
Db 414 YGSTQSGDSAL IAGYSTGTQTAOEGSNLTAGYSTGT-----AGADSSLIAGYSTQSG 469
Qy 301 NHKEVVDSPRSIYTLNDGL-----SDSEAVAVGRYPEDTYNGNPFPLCTIA 347
Db 470 SESSLT IAGYSTGTQTAOEGSNLTAGYSTGTAGADSSLLA-GYSTGT-----S 516
Qy 348 AAEQLYDALY---QMKQGSLEVTDVSLDFKALYSDAATGYSSSSSTYSIVDA--VK 402
Db 517 GSSSLT IAGYSTGTQTAOEGSVLTSGYSTGTQTAAGASNLTGYSTGTAGHESFLIAGYGS 576
Qy 403 TPADEFSVIV-----ETHAASNGSMEQYDKSDGELSARDL-----TWSYALLTA 449
Db 577 TOTAGHKSIL IAGYSTGTQTAOEGSVLT IAGYSTGTAGSSSLIAGYSTGTATAYRMLTA 636
Qy 450 -----NNRNSVVPAMSWGETSASSVPGTCAATSAICTYSSVTWSPSIATG-GTTT 502
Db 637 GYSTGTATAEHSDLVLTGYGSTGTAG-----SNSSLIAGYSTGTATAFKSLITAGYSTGT 691

Qy 503 ATP-----TSGGSVSTIS-----KTTATASKSTTT 528
Db 692 AGERSDLVAGYGS-TSTAGYSSSLIAGYSTGTAGYESTLT 731

RESULT 42
US-10-185-990-11
; Sequence 11, Application US/10185990
; Publication No. US20030073109A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Jae-Gu
; TITLE OF INVENTION: JAE GU PAN ET AL
; FILE REFERENCE: 02589.000100
; CURRENT APPLICATION NUMBER: US/10/185,990
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 1626
; TYPE: PRF
; ORGANISM: Bacillus subtilis
US-10-185-990-11

Query Match 5.6%; Score 155.5; DB 14; Length 1626;
Best Local Similarity 23.6%; Pred. No. 0.00051;
Matches 137; Conservative 66; Mismatches 245; Indels 133; Gaps 27;

Qy 32 SNEATVARTAILNINIGADG-----AMVSGADSGIVVASTDNDP-----YYTWT 77
Db 200 SNETAGNHSDDL IAGYSTGTAGYSTGTQSGEDSSLTAGYSTGTQTAOEGSNLTAGYSTGT 259
Qy 78 --RDGSLVKTIVDLFRNGDTSLSSTIENYISAQAIVOGISNPSGDLSSG-----A 126
Db 260 AGSDSL IAGYSTGTQSGEDSSLT A---GYSTQTQOEGSNLTAGYSTGTACVDSLLA 316
Qy 127 GLGEPEFNDETAYTGSWGRPQRDGPALRATAMIGFQWMLDNGYSTATDIWPLVRND 186
Db 317 GYSTQTQSGDSAL IAGYSTGTQ-----TAQEGSN---LTAGYSTGT-----AGSD 359
Qy 187 LSVVAQYMNQGTGYDLMEVNGSFFITIAV-QHRALVEGSAFATAVGSS-CSMCDQAPET 244
Db 360 SSLLA-----GYGSTQSGDSLSLTAGYSTGTQTAOEGSNLTAGYSTGTAGVDSLLIAG 413
Qy 245 LCVLGSPWTSFPLANFDSRSRSGKDANTL---LGSHTFPBAACDSTFQPC-SPRALA 300
Db 414 YGSTQSGDSAL IAGYSTGTQTAOEGSNLTAGYSTGT-----AGADSSLIAGYSTQSG 469
Qy 301 NHKEVVDSPRSIYTLNDGL-----SDSEAVAVGRYPEDTYNGNPFPLCTIA 347
Db 470 SESSLT IAGYSTGTQTAOEGSNLTAGYSTGTAGADSSLLA-GYSTGT-----S 516
Qy 348 AAEQLYDALY---QMKQGSLEVTDVSLDFKALYSDAATGYSSSSSTYSIVDA--VK 402
Db 517 GSSSLT IAGYSTGTQTAOEGSVLTSGYSTGTQTAAGASNLTGYSTGTAGHESFLIAGYGS 576
Qy 403 TPADEFSVIV-----ETHAASNGSMEQYDKSDGELSARDL-----TWSYALLTA 449
Db 577 TOTAGHKSIL IAGYSTGTQTAOEGSVLT IAGYSTGTAGSSSLIAGYSTGTATAYRMLTA 636
Qy 503 ATP-----TSGGSVSTIS-----KTTATASKSTTT 528
Db 692 AGERSDLVAGYGS-TSTAGYSSSLIAGYSTGTAGYESTLT 731

RESULT 43
US-09-815-242-10932
; Sequence 10932, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

```

1  APPLICANT: Haselbeck, Robert
2  APPLICANT: Ohlsen, Kari L.
3  APPLICANT: Zyckind, Judith W.
4  APPLICANT: Wall, Daniel
5  APPLICANT: Trawick, John D.
6  APPLICANT: Carr, Grant J.
7  APPLICANT: Yamamoto, Robert T.
8  APPLICANT: Xu, H. Howard
9  TITLE OF INVENTION: Identification of Essential Genes in
10 TITLE OF INVENTION: Prokaryotes
11 FILE REFERENCE: ELITRA.011A
12 CURRENT APPLICATION NUMBER: US/09/815,242
13 CURRENT FILING DATE: 2001-03-21
14 PRIOR APPLICATION NUMBER: 60/191,078
15 PRIOR FILING DATE: 2000-03-21
16 PRIOR APPLICATION NUMBER: 60/206,848
17 PRIOR FILING DATE: 2000-05-23
18 PRIOR APPLICATION NUMBER: 60/207,727
19 PRIOR FILING DATE: 2000-05-26
20 PRIOR APPLICATION NUMBER: 60/242,578
21 PRIOR FILING DATE: 2000-10-23
22 PRIOR APPLICATION NUMBER: 60/253,625
23 PRIOR FILING DATE: 2000-11-27
24 PRIOR APPLICATION NUMBER: 60/257,931
25 PRIOR FILING DATE: 2000-12-22
26 PRIOR APPLICATION NUMBER: 60/269,308
27 PRIOR FILING DATE: 2001-02-16
28 NUMBER OF SEQ ID NOS: 14110
29 SOFTWARE: FastSeq for Windows Version 4.0
30 SEQ ID NO 10932
31 LENGTH: 1031
32 TYPE: prt
33 ORGANISM: Enterococcus faecalis
34 US-09-815-242-10932

```

```

DB              874 STSSTSETSNTNESNTPSTSKTSSTSESSAS 906

RESULT 44
US-10-120-801-63
; Sequence 63, Application US/10120801
; Publication No. US20030203843A1
GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-340
CURRENT APPLICATION NUMBER: US/10/120,801
CURRENT FILING DATE: 2002-04-11
PRIOR APPLICATION NUMBER: 60/285748
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: 60/286068
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 60/286292
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/286334
PRIOR FILING DATE: 2001-05-03
PRIOR APPLICATION NUMBER: 60/291241
PRIOR FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 60/322284
PRIOR FILING DATE: 2001-09-14
PRIOR APPLICATION NUMBER: 60/285609
PRIOR FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 155
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 63
LENGTH: 2117
TYPE: PRT
ORGANISM: Neuraospora crassa
US-10-120-801-63

```

Query Match	5.6%	Score 154.5	DB 9	Length 1031
Best Local Similarity	21.2%	Pred. No. 0.00031		
Matches 109	Conservative 69	Mismatches 206	Indels 129	Gaps 18

QY	75	TWTRDSGLVLTLYDVFRRNGDTSLSTIENYISAGAIYQGISNPSGDSLSGAG-----	127
DB	469	TW-----ESLSPFLVEQYKNTPIPVANTLATPV-ATTIVQGGSIDPQVLVANTNDADAVP	524
QY	128	-----LGPFRNVDETAATGSGWRQGRDPALRATAMIGPQMLDNGYSTA	175
DB	525	AAKIVNPEVLNIEIGQAKVQVLTFTETGS-----LVTVNVPTV	562
QY	176	TDIVMPLVRNDLSVYAQYV-----NQTGYDLMEEVNG-----SSEFTIAVQHRALVE	222
DB	563	ED-----PAKKLDLSQGLQYVYASIGENQYTVYSSKQLFPAIGPKTIVPSYQQATO--LAAE	617
QY	223	GSAFATAVGSSCCSCSDQAPRILCYLQSGFWTGSFTLANPDSRSKGDANTLGSIHTPDP	282
DB	618	GVV-----SGDKTQEQVDOLTSMLQGS--AMKVLVKKRADITLIERAEKENEIASYHKUDE	668
QY	283	EACDDSTFGQCSPPALANHKKEVNDSPRSIYTLANG-----LADSEAV-AVGRY	330
DB	669	SVYTTDSV-----QMQEALDIT-----TTGEGSSKQLOQLLAWDEBELLEPTLIGF	715
QY	331	PEDTYNNGPWFELCTIAAAEQULYDALYOMDKQGSLEVNDVSLDFPKALYSDAATGYS	390
DB	716	K-----TPDAQKRINQLTGITIKTALLLLEKSTETTSNTSSESTSTTSES	762
QY	391	SSTYSIYDAVKTFPDAVGVSIVETHAANGSNGSEQYDKSDGQLSARDLTMSYALLTYAN	450
DB	763	SSTSESSSTSTTN-----ESSSTSESSSTSTTNSSSSTSESSSTSTSTSESSTSES	813
QY	451	NRNRNV-VPASWGETSASSVPQTCATGATGALIGTSSVTVSWPISYATGCTTTATPTGSG	509
DB	814	STPSTSSSSTSESSSTSTTNSSSSTSESSSTPSTTSSSSTSTSESSSTSTSTSESSTSES	873
QY	510	SVTST-----SKTATLAKSTISTTTRSGMS	533

```
Db      1072 AVVGDWEYAGCRTDNASPATRALNDRIYQGSSTNTIESCAQACAGFSFFGVEMGEYC 1131
Qy      315 LNDGLSDSEAVAVNGRYPEDT---YNGNPMFLCTLAAEQLYDLYQMDQSLLEYTDV 371
Db      1132 GNN-LNPGSTVA---DEKTCNVCGGDPTELCGGSGRISVY-----KQKGTIVGNPS 1179
Qy      372 LDFPKALYSDAATGTYSSTYSIYDAVKTPAD-----GFSIVETHAANGSMSEQ 425
Db      1180 TGVSS--SGTASGTASATASASTSSSAIASGTPGNPOSIGQYSSLCYSDAVASRLQ 1237
Qy      426 YKSDGEOLASARDLT-----WSY-----AALLTANRRNSVVPASWGETSA---SSVPG 471
Db      1238 GKNTGSNWMSLDDCATYCAGYKFGTEYSAECFCGNDLNGAAPVTDGRCNMLCNGNQO 1297
Qy      472 TCAATSAITGYSVYVTSWPSIVATGTTTATPTGSGSVTSTSKTTATASKTST 526
Db      1298 ICGSGNGLSMYQLNPNGTSSSVYASGSATQASATASGTASGTAASSSTALATTSS 1352
```

RESULT 45

```
US-10-228-063-53
; Sequence 53, Application US/10228063
; Publication No. US20030135885A1
; GENERAL INFORMATION:
; APPLICANT: Lanahan, Mike
; TITLE OF INVENTION: Self-processing Plants and Plant Parts
; FILE REFERENCE: 109846.317
; CURRENT APPLICATION NUMBER: US/10/228,063
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-228-063-53
```

```
Query Match      5.5%; Score 152; DB 14; Length 45;
Best Local Similarity 88.9%; Pred. No. 4.5e-06;
Matches 32; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy      495 ATGGTTTATPTGSGSVTSTSKTTATASKTSTTRS 530
Db      1 ATGGTTTATPTGSGSVTSTSKTTATASKTSTTSS 36
```

Search completed: June 28, 2004, 07:50:52
Job time : 54 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:42:07 ; Search time 23 Seconds
(without alignments)
1198.620 Million cell updates/sec

Title: US-10-038-723-2

Perfect score: 2771
Sequence: 1 MSFRSLALSGVCTGLANV.....SKTTATSKSTTTRGMSL 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2771	100.0	534	3	US-09-198-672-2
2	2771	100.0	534	4	US-09-199-290-9
3	2771	100.0	534	4	US-09-632-392-2
4	2771	100.0	534	4	US-09-455-679-1
5	2771	100.0	534	4	US-09-351-814-2
6	2771	100.0	534	4	US-09-821-616-9
7	2742	99.0	640	4	US-09-351-814-13
8	2628	94.8	616	4	US-09-236-063-1
9	1689	61.0	618	3	US-09-199-290-34
10	1689	61.0	618	4	US-09-821-616-34
11	1674.5	60.4	591	3	US-09-199-290-7
12	1674.5	60.4	591	4	US-09-821-616-7
13	1615.5	58.3	630	4	US-09-704-419-2
14	1577	56.9	626	2	US-08-596-300A-7
15	1577	56.9	626	2	US-08-596-300A-14
16	1425.5	51.4	581	4	US-09-534-407-3
17	1425.5	51.4	581	4	US-09-999-201B-4
18	1321	47.7	616	1	US-08-385-370-2
19	1321	47.7	616	1	US-08-385-370-4
20	589	21.3	806	1	US-08-270-076A-11
21	160	5.8	35	3	US-09-199-290-4
22	160	5.8	35	4	US-09-821-616-4
23	156.5	5.6	2137	4	US-09-134-001C-4463
24	142	5.1	536	4	US-09-463-712C-10
25	136	4.9	1026	2	US-08-614-377A-7
26	136	4.9	1026	3	US-09-142-648B-7
27	133	4.8	1529	2	US-08-728-470-10

28	133	4.8	1529	3	US-08-719-641-10	Sequence 10, Appl
29	132	4.8	1426	3	US-09-136-574A-43	Sequence 43, Appl
30	130.5	4.7	1537	1	US-08-325-267A-2	Sequence 2, Appl
31	130	4.7	1600	2	US-08-617-697-10	Sequence 10, Appl
32	129	4.7	1026	1	US-08-194-290-7	Sequence 7, Appl
33	128.5	4.6	673	4	US-09-107-532A-5134	Sequence 5134, Ap
34	128	4.6	1721	3	US-08-700-651-5	Sequence 5, Appl
35	128	4.6	1721	3	US-08-928-361B-6	Sequence 6, Appl
36	128	4.6	1721	4	US-09-588-995A-6	Sequence 6, Appl
37	127.5	4.6	751	6	5457037-5	Patent No. 5457037
38	127.5	4.6	776	3	US-09-346-237-7	Sequence 7, Appl
39	127	4.6	630	2	US-08-797-366-3	Sequence 3, Appl
40	127	4.6	630	2	US-08-956-268-3	Sequence 3, Appl
41	127	4.6	906	4	US-09-717-364A-15	Sequence 15, Appl
42	126.5	4.6	776	3	US-09-346-237-4	Sequence 4, Appl
43	126	4.5	288	4	US-09-216-393B-341	Sequence 341, App
44	126	4.5	288	4	US-09-216-393B-344	Sequence 344, App
45	124.5	4.5	894	3	US-08-362-525-22	Sequence 22, Appl
46	124.5	4.5	894	3	US-08-971-692-15	Sequence 15, Appl
47	124.5	4.5	3892	4	US-09-328-352-5503	Sequence 5503, Ap
48	123	4.4	562	6	5258502-2	Patent No. 5258502
49	123	4.4	725	2	US-08-816-105A-1	Sequence 1, Appl
50	122.5	4.4	750	6	5457037-3	Patent No. 5457037
51	122	4.4	862	1	US-08-325-267A-4	Sequence 4, Appl
52	121.5	4.4	2411	4	US-09-268-347-36	Sequence 36, Appl
53	121	4.4	629	4	US-09-241-581B-6	Sequence 6, Appl
54	121	4.4	629	4	US-08-265-428-6	Sequence 6, Appl
55	121	4.4	629	5	PCT-US95-07721-6	Sequence 567, Ap
56	120.5	4.3	823	4	US-09-107-532A-5667	Sequence 9, Appl
57	120	4.3	1338	2	US-08-728-470-9	Sequence 9, Appl
58	120	4.3	1338	3	US-08-719-641-9	Sequence 9, Appl
59	120	4.3	1599	2	US-08-617-697-9	Sequence 9, Appl
60	119.5	4.3	669	4	US-09-107-532A-6532	Sequence 6532, Ap
61	119	4.3	693	1	US-08-056-200-111	Sequence 111, App
62	119	4.3	693	2	US-08-800-644-111	Sequence 111, App
63	119	4.3	2736	2	US-09-252-991A-30227	Sequence 30227, A
64	118	4.3	52	6	5198345-14	Patent No. 5198345
65	117.5	4.2	1912	1	US-08-409-995-4	Sequence 4, Appl
66	117.5	4.2	1912	3	US-08-685-467-4	Sequence 4, Appl
67	117.5	4.2	2039	4	US-09-077-098A-7	Sequence 7, Appl
68	117.5	4.2	2353	3	US-09-377-155-33	Sequence 33, Appl
69	117.5	4.2	2353	3	US-08-913-942-4	Sequence 4, Appl
70	117.5	4.2	2353	4	US-09-669-974-33	Sequence 33, Appl
71	117.5	4.2	2353	4	US-09-797-862-33	Sequence 33, Appl
72	117.5	4.2	2354	4	US-09-268-347-47	Sequence 47, Appl
73	117	4.2	1004	4	US-09-206-942-57	Sequence 57, Appl
74	117	4.2	1010	4	US-09-206-942-55	Sequence 55, Appl
75	116	4.2	413	1	US-08-295-676A-2	Sequence 2, Appl
76	116	4.2	413	1	US-08-750-077-6	Sequence 6, Appl
77	116	4.2	1388	4	US-08-948-591-2	Sequence 2, Appl
78	115.5	4.2	270	4	US-09-543-681A-5641	Sequence 5641, Ap
79	115	4.2	657	4	US-09-134-000C-3553	Sequence 3553, Ap
80	115	4.2	617	1	US-09-543-681A-7109	Sequence 7109, Ap
81	114.5	4.1	617	1	US-08-361-920-29	Sequence 29, Appl
82	114.5	4.1	617	1	US-08-479-939-29	Sequence 29, Appl
83	114.5	4.1	617	1	US-08-483-432-29	Sequence 29, Appl
84	114.5	4.1	1751	3	US-09-136-574A-44	Sequence 44, Appl
85	114	4.1	24	6	5198345-26	Patent No. 5198345
86	113.5	4.1	550	4	US-09-198-452A-613	Sequence 613, Appl
87	113.5	4.1	573	3	US-08-881-742-2	Sequence 2, Appl
88	113	4.1	28	6	5198345-19	Patent No. 5198345
89	113	4.1	553	1	US-08-651-572-2	Sequence 2, Appl
90	113	4.1	553	3	US-09-066-544-2	Sequence 2, Appl
91	113	4.1	553	3	US-08-951-086-2	Sequence 2, Appl
92	113	4.1	553	3	US-09-430-669-2	Sequence 2, Appl
93	113	4.1	715	4	US-09-620-412C-321	Sequence 321, App
94	113	4.1	715	4	US-09-598-419-321	Sequence 321, App
95	113	4.1	1060	4	US-08-911-393-2	Sequence 2, Appl
96	112.5	4.1	915	4	US-09-206-942-35	Sequence 35, Appl
97	112.5	4.1	1222	4	US-09-206-942-37	Sequence 37, Appl
98	112.5	4.1	1228	4	US-09-206-942-34	Sequence 34, Appl
99	112.5	4.1	1837	3	US-08-928-361B-5	Sequence 5, Appl
100	112.5	4.1	1837	4	US-09-588-995A-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1

```
US-09-198-672-2
; Sequence 2, Application US/09198672
; Patent No. 6129788
; GENERAL INFORMATION:
; APPLICANT: Liaw, Gin
; APPLICANT: Pedersen, Sven
; APPLICANT: Hendiksen, Sven
; TITLE OF INVENTION: A Method of Producing Saccharide
; FILE REFERENCE: 5318.200-US
; CURRENT APPLICATION NUMBER: US/09/198, 672
; CURRENT FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 534
; TYPE: PR1
; ORGANISM: Aspergillus Niger
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-198-672-2
```

Query Match 100.0%; Score 2771; DB 3; Length 534;

Best Local Similarity 100.0%; Pred. No. 1.1e-231; Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
QY 61 VVASPSTNDPDYFYTWTRDSGLVLTIVLFRNGDTSLSTIENYISAQIIVOGISNPSG 120
DB 61 VVASPSTNDPDYFYTWTRDSGLVLTIVLFRNGDTSLSTIENYISAQIIVOGISNPSG 120
QY 121 DISSGAGLGEPRKRVNDETATYTGSGRPQRDGPALRAMIGFGOMLLDNGYSTATDIW 180
DB 121 DISSGAGLGEPRKRVNDETATYTGSGRPQRDGPALRAMIGFGOMLLDNGYSTATDIW 180
QY 121 DISSGAGLGEPRKRVNDETATYTGSGRPQRDGPALRAMIGFGOMLLDNGYSTATDIW 180
DB 121 DISSGAGLGEPRKRVNDETATYTGSGRPQRDGPALRAMIGFGOMLLDNGYSTATDIW 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNCGSFETIAVOHRALVEGSAFATAVAGSSCWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNCGSFETIAVOHRALVEGSAFATAVAGSSCWCDSQ 240
QY 241 APEILCYLQSFMTGSLILANFDSRSGKDNATLLGSIHTFDPBAACDDSTFQPCSPRALA 300
DB 241 APEILCYLQSFMTGSLILANFDSRSGKDNATLLGSIHTFDPBAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAEOQLDALYQMD 360
DB 301 NHKEVVDSPRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAEOQLDALYQMD 360
QY 361 KQSGLEVTDVSLDFFKALYSDAATGTSSSSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
DB 361 KQSGLEVTDVSLDFFKALYSDAATGTSSSSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
QY 421 SMEQYDKSDGEOLASRDLTWSYAALLTANNRNSVVPASWGETSASVPGCTCAATSATG 480
DB 421 SMEQYDKSDGEOLASRDLTWSYAALLTANNRNSVVPASWGETSASVPGCTCAATSATG 480
QY 481 TYSSTVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTRSGMSL 534
DB 481 TYSSTVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTRSGMSL 534
```

RESULT 2

US-09-199-290-9
; Sequence 9, Application US/09199290

; Patent No. 6255084

; GENERAL INFORMATION:

; APPLICANT: Nielsen, Bjørne R.

; APPLICANT: Nielsen, Ruby

; APPLICANT: Lehmbeck, Jan

; TITLE OF INVENTION: Thermostable Glucoamylase

; FILE REFERENCE: 5279.200-US

; CURRENT APPLICATION NUMBER: US/09/199, 290

; EARLIER FILING DATE: 1998-11-24

; EARLIER FILING DATE: 1997-12-30

; EARLIER FILING DATE: 1997-12-30

; EARLIER FILING DATE: 1998-07-10

; EARLIER FILING DATE: 1998-07-10

; EARLIER FILING DATE: 1998-01-08

; EARLIER FILING DATE: 1998-01-08

; EARLIER FILING DATE: 1997-11-26

; EARLIER FILING DATE: 1997-11-26

; EARLIER FILING DATE: 1998-06-30

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 9

; LENGTH: 534

; TYPE: PR1

; ORGANISM: Aspergillus niger

; FEATURE:

; NAME/KEY: SIGNAL

; LOCATION: (1)...(24)

Query Match 100.0%; Score 2771; DB 3; Length 534;

Best Local Similarity 100.0%; Pred. No. 1.1e-231; Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
QY 61 VVASPSTNDPDYFYTWTRDSGLVLTIVLFRNGDTSLSTIENYISAQIIVOGISNPSG 120
DB 61 VVASPSTNDPDYFYTWTRDSGLVLTIVLFRNGDTSLSTIENYISAQIIVOGISNPSG 120
QY 121 DISSGAGLGEPRKRVNDETATYTGSGRPQRDGPALRAMIGFGOMLLDNGYSTATDIW 180
DB 121 DISSGAGLGEPRKRVNDETATYTGSGRPQRDGPALRAMIGFGOMLLDNGYSTATDIW 180
QY 121 DISSGAGLGEPRKRVNDETATYTGSGRPQRDGPALRAMIGFGOMLLDNGYSTATDIW 180
DB 121 DISSGAGLGEPRKRVNDETATYTGSGRPQRDGPALRAMIGFGOMLLDNGYSTATDIW 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNCGSFETIAVOHRALVEGSAFATAVAGSSCWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNCGSFETIAVOHRALVEGSAFATAVAGSSCWCDSQ 240
QY 241 APEILCYLQSFMTGSLILANFDSRSGKDNATLLGSIHTFDPBAACDDSTFQPCSPRALA 300
DB 241 APEILCYLQSFMTGSLILANFDSRSGKDNATLLGSIHTFDPBAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAEOQLDALYQMD 360
DB 301 NHKEVVDSPRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAEOQLDALYQMD 360
QY 361 KQSGLEVTDVSLDFFKALYSDAATGTSSSSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
DB 361 KQSGLEVTDVSLDFFKALYSDAATGTSSSSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
QY 421 SMEQYDKSDGEOLASRDLTWSYAALLTANNRNSVVPASWGETSASVPGCTCAATSATG 480
DB 421 SMEQYDKSDGEOLASRDLTWSYAALLTANNRNSVVPASWGETSASVPGCTCAATSATG 480
QY 481 TYSSTVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTRSGMSL 534
DB 481 TYSSTVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTRSGMSL 534
```



```
RESULT 3
US-09-632-392-2
; Sequence 2, Application US/09632392
; Patent No. 6303346
; GENERAL INFORMATION:
; APPLICANT: Liaw, Gin
; APPLICANT: Pedersen, Sven
; APPLICANT: Hendriksen, Sven
; TITLE OF INVENTION: A Method of Producing Saccharide
; FILE REFERENCE: 5318.200-US
; CURRENT APPLICATION NUMBER: US/09/632,392
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 09/198,672
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FaastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus Niger
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-632-392-2

Query Match      100.0%; Score 2771; DB 4; Length 534;
Best Local Similarity 100.0%; Pred. No. 1,1e-231;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNIGADGAWVSGADSGI 60
DB 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNIGADGAWVSGADSGI 60
QY 61 VVASPSTNDPDYFYTTTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGINSNPSG 120
DB 61 VVASPSTNDPDYFYTTTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGINSNPSG 120
QY 121 DISSGAGLGEPRKPNVDETAYTSGWGRPQRDPALRATAMIGFQWLLDNGYSTANDIYW 180
DB 121 DISSGAGLGEPRKPNVDETAYTSGWGRPQRDPALRATAMIGFQWLLDNGYSTANDIYW 180
QY 121 DISSGAGLGEPRKPNVDETAYTSGWGRPQRDPALRATAMIGFQWLLDNGYSTANDIYW 180
DB 121 DISSGAGLGEPRKPNVDETAYTSGWGRPQRDPALRATAMIGFQWLLDNGYSTANDIYW 180
QY 181 PLVRNDLSVVAQYWNQGYDLMEEVNCGSFPTIAYOHRALVEGSARFATAVGSSCWCDSDQ 240
DB 181 PLVRNDLSVVAQYWNQGYDLMEEVNCGSFPTIAYOHRALVEGSARFATAVGSSCWCDSDQ 240
QY 241 APEILCYLOSFWTGSFILANFDSRSRGKDANTLLGSIHTFDPRAACDDSTFQPCSPRALA 300
DB 241 APEILCYLOSFWTGSFILANFDSRSRGKDANTLLGSIHTFDPRAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSTRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPFICTLAAAEQLYDALYQMD 360
DB 301 NHKEVVDSTRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPFICTLAAAEQLYDALYQMD 360
QY 301 NHKEVVDSTRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPFICTLAAAEQLYDALYQMD 360
DB 301 NHKEVVDSTRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPFICTLAAAEQLYDALYQMD 360
QY 361 KQGSLEVTDVSLDFEFALYSDAATGTYSSTSSSTYSISVDAVKTFADGPFVSIETHAASNG 420
DB 361 KQGSLEVTDVSLDFEFALYSDAATGTYSSTSSSTYSISVDAVKTFADGPFVSIETHAASNG 420
QY 421 SMSEQYDKSDGEOLSARDLTWYSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOLSARDLTWYSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TYSSVTVTWPSPISVATGCTTTTATPTGSGSVSTSKTTATASKTSTTTTSSGMSL 534
DB 481 TYSSVTVTWPSPISVATGCTTTTATPTGSGSVSTSKTTATASKTSTTTTSSGMSL 534

RESULT 4
US-09-455-679-1
; Sequence 1, Application US/09455679
; Patent No. 6329186
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne R.
```

```
; APPLICANT: Svendsen, Allan
; APPLICANT: Bojsen, Kirsten
; APPLICANT: Vind, Jesper
; APPLICANT: Pedersen, Henrik
; TITLE OF INVENTION: Glucoamyloses With N-Terminal Extensions
; FILE REFERENCE: 5691.200-US
; CURRENT APPLICATION NUMBER: US/09/455,679
; PRIOR FILING DATE: 1999-12-07
; EARLIER APPLICATION NUMBER: PA 1998 01616
; EARLIER FILING DATE: 1998-12-07
; EARLIER APPLICATION NUMBER: PA 1999 00409
; EARLIER FILING DATE: 1999-03-24
; EARLIER APPLICATION NUMBER: 60/111,674
; EARLIER FILING DATE: 1998-12-10
; EARLIER APPLICATION NUMBER: 60/126,740
; EARLIER FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus niger
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-455-679-1

Query Match      100.0%; Score 2771; DB 4; Length 534;
Best Local Similarity 100.0%; Pred. No. 1,1e-231;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNIGADGAWVSGADSGI 60
DB 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNIGADGAWVSGADSGI 60
QY 61 VVASPSTNDPDYFYTTTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGINSNPSG 120
DB 61 VVASPSTNDPDYFYTTTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGINSNPSG 120
QY 121 DISSGAGLGEPRKPNVDETAYTSGWGRPQRDPALRATAMIGFQWLLDNGYSTANDIYW 180
DB 121 DISSGAGLGEPRKPNVDETAYTSGWGRPQRDPALRATAMIGFQWLLDNGYSTANDIYW 180
QY 121 DISSGAGLGEPRKPNVDETAYTSGWGRPQRDPALRATAMIGFQWLLDNGYSTANDIYW 180
DB 121 DISSGAGLGEPRKPNVDETAYTSGWGRPQRDPALRATAMIGFQWLLDNGYSTANDIYW 180
QY 181 PLVRNDLSVVAQYWNQGYDLMEEVNCGSFPTIAYOHRALVEGSARFATAVGSSCWCDSDQ 240
DB 181 PLVRNDLSVVAQYWNQGYDLMEEVNCGSFPTIAYOHRALVEGSARFATAVGSSCWCDSDQ 240
QY 241 APEILCYLOSFWTGSFILANFDSRSRGKDANTLLGSIHTFDPRAACDDSTFQPCSPRALA 300
DB 241 APEILCYLOSFWTGSFILANFDSRSRGKDANTLLGSIHTFDPRAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSTRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPFICTLAAAEQLYDALYQMD 360
DB 301 NHKEVVDSTRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPFICTLAAAEQLYDALYQMD 360
QY 301 NHKEVVDSTRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPFICTLAAAEQLYDALYQMD 360
DB 301 NHKEVVDSTRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPFICTLAAAEQLYDALYQMD 360
QY 361 KQGSLEVTDVSLDFEFALYSDAATGTYSSTSSSTYSISVDAVKTFADGPFVSIETHAASNG 420
DB 361 KQGSLEVTDVSLDFEFALYSDAATGTYSSTSSSTYSISVDAVKTFADGPFVSIETHAASNG 420
QY 421 SMSEQYDKSDGEOLSARDLTWYSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOLSARDLTWYSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TYSSVTVTWPSPISVATGCTTTTATPTGSGSVSTSKTTATASKTSTTTTSSGMSL 534
DB 481 TYSSVTVTWPSPISVATGCTTTTATPTGSGSVSTSKTTATASKTSTTTTSSGMSL 534

RESULT 5
US-09-351-814-2
; Sequence 2, Application US/09351814
; Patent No. 6352851
; GENERAL INFORMATION:
```

```
/ APPLICANT: Nielsen, Bjarne Roenfeldt
/ APPLICANT: Svendsen, Allan
/ APPLICANT: Pedersen, Henrik
/ APPLICANT: Vind, Jesper
/ APPLICANT: Hendriksen, Hanne Vang
/ APPLICANT: Frandsen, Torben Peter
/ TITLE OF INVENTION: Glucoamylase Variants
/ FILE REFERENCE: 5636.200-US
/ CURRENT APPLICATION NUMBER: US/09/351,814
/ EARLIER FILING DATE: 1999-07-12
/ EARLIER APPLICATION NUMBER: PA 1998 00937
/ EARLIER FILING DATE: 1998-07-15
/ EARLIER APPLICATION NUMBER: PA 1998 01667
/ EARLIER FILING DATE: 1998-12-17
/ EARLIER APPLICATION NUMBER: 60/093,528
/ EARLIER FILING DATE: 1998-07-21
/ EARLIER APPLICATION NUMBER: 60/115,545
/ EARLIER FILING DATE: 1999-01-12
/ NUMBER OF SEQ. ID NOS: 81
/ SOFTWARE: FaetSeq for Windows Version 3.0
/ SEQ. ID NO 2
/ LENGTH: 534
/ TYPE: PR1
/ ORGANISM: Aspergillus niger
/ FEATURE:
/ NAME/KEY: SIGNAL
/ LOCATION: (1)...(24)
US-09-351-814-2
```

Query Match 100.0%; Score 2771; DB 4; Length 534;

Best Local Similarity 100.0%; Pred. No. 1.1e-231; Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTALINNIAGDAMVSGADSGI 60
Db 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTALINNIAGDAMVSGADSGI 60
Qy 61 VVASPSTNDPDIYFTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYVGISNPSG 120
Db 61 VVASPSTNDPDIYFTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYVGISNPSG 120
Qy 121 DLSGAGLGPKNVDATAVYTGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Db 121 DLSGAGLGPKNVDATAVYTGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Qy 181 PLVRNDLSVYAOYWNQGYDLMEEVNSSFPTIAVQHRALVEGSARFATVAGSSCWCDSQ 240
Db 181 PLVRNDLSVYAOYWNQGYDLMEEVNSSFPTIAVQHRALVEGSARFATVAGSSCWCDSQ 240
Qy 241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDSTFQPCSPRALA 300
Db 241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDSTFQPCSPRALA 300
Qy 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTTAAAEQLDALYQMD 360
Db 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTTAAAEQLDALYQMD 360
Qy 361 KQGSLEVTIVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTFADGPFVSIIVETHAASNG 420
Db 361 KQGSLEVTIVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTFADGPFVSIIVETHAASNG 420
Qy 421 SMEQYDKSDGEOLSARDLTWSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
Db 421 SMEQYDKSDGEOLSARDLTWSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
Qy 481 TYSSVTVTSWPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTRSGMSL 534
Db 481 TYSSVTVTSWPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTRSGMSL 534
```

RESULT 6
US-09-821-616-9
; Sequence 9, Application US/09821616

```
/ Patent No. 6620924
/ GENERAL INFORMATION:
/ APPLICANT: Nielsen, Bjarne R.
/ APPLICANT: Nielsen, Rudy
/ APPLICANT: Lehmbeck, Jan
/ TITLE OF INVENTION: Thermostable Glucoamylase
/ FILE REFERENCE: 5279.200-US
/ CURRENT APPLICATION NUMBER: US/09/821,616
/ EARLIER FILING DATE: 2001-03-29
/ EARLIER APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/199,290
/ EARLIER FILING DATE: 1998-11-24
/ EARLIER APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1557/97
/ EARLIER FILING DATE: 1997-12-30
/ EARLIER APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 0925/98
/ EARLIER FILING DATE: 1998-07-10
/ EARLIER APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,746
/ EARLIER FILING DATE: 1998-01-08
/ EARLIER APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/094,344
/ EARLIER FILING DATE: 1998-07-28
/ EARLIER APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/979,673
/ EARLIER FILING DATE: 1997-11-26
/ EARLIER APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/107,657
/ EARLIER FILING DATE: 1998-06-30
/ NUMBER OF SEQ. ID NOS: 34
/ SOFTWARE: FaetSeq for Windows Version 3.0
/ SEQ. ID NO 9
/ LENGTH: 534
/ TYPE: PR1
/ ORGANISM: Aspergillus niger
/ FEATURE:
/ NAME/KEY: SIGNAL
/ LOCATION: (1)...(24)
US-09-821-616-9
```

Query Match 100.0%; Score 2771; DB 4; Length 534;

Best Local Similarity 100.0%; Pred. No. 1.1e-231; Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTALINNIAGDAMVSGADSGI 60
Db 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTALINNIAGDAMVSGADSGI 60
Qy 61 VVASPSTNDPDIYFTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYVGISNPSG 120
Db 61 VVASPSTNDPDIYFTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYVGISNPSG 120
Qy 121 DLSGAGLGPKNVDATAVYTGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Db 121 DLSGAGLGPKNVDATAVYTGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Qy 181 PLVRNDLSVYAOYWNQGYDLMEEVNSSFPTIAVQHRALVEGSARFATVAGSSCWCDSQ 240
Db 181 PLVRNDLSVYAOYWNQGYDLMEEVNSSFPTIAVQHRALVEGSARFATVAGSSCWCDSQ 240
Qy 241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDSTFQPCSPRALA 300
Db 241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDSTFQPCSPRALA 300
Qy 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTTAAAEQLDALYQMD 360
Db 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTTAAAEQLDALYQMD 360
Qy 361 KQGSLEVTIVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTFADGPFVSIIVETHAASNG 420
Db 361 KQGSLEVTIVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTFADGPFVSIIVETHAASNG 420
Qy 421 SMEQYDKSDGEOLSARDLTWSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
Db 421 SMEQYDKSDGEOLSARDLTWSYALLLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
Qy 481 TYSSVTVTSWPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTRSGMSL 534
Db 481 TYSSVTVTSWPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTRSGMSL 534
```

RESULT 7

US-09-351-814-13
Sequence 13, Application US/09351814
Patent No. 6352851
GENERAL INFORMATION:
APPLICANT: Nielsen, Bjarne Roenfeldt
APPLICANT: Svendsen, Allan
APPLICANT: Pedersen, Henrik
APPLICANT: Vind, Jesper
APPLICANT: Hendiksen, Hanne Vang
APPLICANT: Frandsen, Torben Peter
TITLE OF INVENTION: Glucoamylase Variants
FILE REFERENCE: 5636-200-US
CURRENT APPLICATION NUMBER: US/09/351, 814
CURRENT FILING DATE: 1999-07-12
EARLIER APPLICATION NUMBER: PA 1998 00937
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: PA 1998 01667
EARLIER FILING DATE: 1998-12-17
EARLIER APPLICATION NUMBER: 60/093, 528
EARLIER FILING DATE: 1998-07-21
EARLIER APPLICATION NUMBER: 60/115, 545
EARLIER FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 81
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 13
LENGTH: 640
TYPE: PR1
ORGANISM: ASPERGILLUS NIGER
US-09-351-814-13

Query Match 99.0%; Score 2742; DB 4; Length 640;
Best Local Similarity 99.6%; Pred. No. 4, 9e-229;
Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 MSFRLSLALSGVCTGTLANVYSKRATLDSWLSNEATVAPRAILNINIGADGAWSGADSGI 60
1 MSFRLSLALSGVCTGTLANVYSKRATLDSWLSNEATVAPRAILNINIGADGAWSGADSGI 60
61 VVASPSTNDPNDYFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAVOGISNPSG 120
61 VVASPSTNDPNDYFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAVOGISNPSG 120
121 DISSGAGLGEPRKNNVDETATVSGWRPQRDPALRATAMIGFQWLLDNGYSTADIYW 180
121 DISSGAGLGEPRKNNVDETATVSGWRPQRDPALRATAMIGFQWLLDNGYSTADIYW 180
181 PLVRNDLSVVAQVWNGTGYDLMEVNGSSFFTLAVOHRALVEGSAFPAVAVGSSCSWCDQ 240
181 PLVRNDLSVVAQVWNGTGYDLMEVNGSSFFTLAVOHRALVEGSAFPAVAVGSSCSWCDQ 240
241 APEILCYLOSFWTGSFLLANFDSRSRSGKANTLGSITHTDPAAACDSTFQCSPRALA 300
241 APEILCYLOSFWTGSFLLANFDSRSRSGKANTLGSITHTDPAAACDSTFQCSPRALA 300
301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLYDALYQWD 360
301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLYDALYQWD 360
361 KQSSLEVTDSVLPFFALVSDAATGYSSSSSTYSSIVDAVKTFAQGFVSIVETHAASNG 420
361 KQSSLEVTDSVLPFFALVSDAATGYSSSSSTYSSIVDAVKTFAQGFVSIVETHAASNG 420
421 SMSEQYDKSDGDELSARDLTWSYAALLTANRRNSVVPASGSETASASVGTCAATSAIG 480
421 SMSEQYDKSDGDELSARDLTWSYAALLTANRRNSVVPASGSETASASVGTCAATSAIG 480
481 TVSSVTVTSWPSIVATGGTTTATPTGSGSVTSTSKTTATASKTSTTRS 530
481 TVSSVTVTSWPSIVATGGTTTATPTGSGSVTSTSKTTATASKTSTTRS 530

RESULT 8

US-09-236-063-1
Sequence 1, Application US/09236063
Patent No. 6537792
GENERAL INFORMATION:
APPLICANT: Allen, Martin
APPLICANT: Pang, Tsuei-Yun
APPLICANT: Li, Yuxing
APPLICANT: Liu, Hsuan-Liang
APPLICANT: Chen, Hsui-Mei
APPLICANT: Coutinho, Pedro
APPLICANT: Hanzacko, Richard
APPLICANT: Ford, Clark
TITLE OF INVENTION: PROTEIN ENGINEERING OF GLUCOAMYLASE TO
TITLE OF INVENTION: INCREASE PH OPTIMUM, SUBSTRATE SPECIFICITY AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 6537792thwestern Hwy.
CITY: Farmington Hills
STATE: Michigan
COUNTRY: US
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/236, 063
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: 0812.00001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 539-5055
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Aspergillus
US-09-236-063-1

Query Match 94.8%; Score 2628; DB 4; Length 616;
Best Local Similarity 99.6%; Pred. No. 3, 5e-219;
Matches 504; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

25 ATLDLSWLSNEATVAPRAILNINIGADGAWSGADSGIVVASPSTNDPNDYFYTWTRDSGLV 84
25 ATLDLSWLSNEATVAPRAILNINIGADGAWSGADSGIVVASPSTNDPNDYFYTWTRDSGLV 84
1 ATLDLSWLSNEATVAPRAILNINIGADGAWSGADSGIVVASPSTNDPNDYFYTWTRDSGLV 60
1 ATLDLSWLSNEATVAPRAILNINIGADGAWSGADSGIVVASPSTNDPNDYFYTWTRDSGLV 60
85 KTLVDFPRNGDTSLSSTIENYISAQAVOGISNPSGDLSSGAGLGEPRKNNVDETATVSGW 144
85 KTLVDFPRNGDTSLSSTIENYISAQAVOGISNPSGDLSSGAGLGEPRKNNVDETATVSGW 144
61 KTLVDFPRNGDTSLSSTIENYISAQAVOGISNPSGDLSSGAGLGEPRKNNVDETATVSGW 120
61 KTLVDFPRNGDTSLSSTIENYISAQAVOGISNPSGDLSSGAGLGEPRKNNVDETATVSGW 120
145 GRPQRDPALRATAMIGFQWLLDNGYSTADIWPLVRNDLSVVAQVWNGTGYDLMEZ 204
145 GRPQRDPALRATAMIGFQWLLDNGYSTADIWPLVRNDLSVVAQVWNGTGYDLMEZ 204
121 GRPQRDPALRATAMIGFQWLLDNGYSTADIWPLVRNDLSVVAQVWNGTGYDLMEZ 180
121 GRPQRDPALRATAMIGFQWLLDNGYSTADIWPLVRNDLSVVAQVWNGTGYDLMEZ 180
205 VNGSSFFTLAVOHRALVEGSAFPAVAVGSSCSWCDQAPPEILCYLOSFWTGSFLLANFDS 264
205 VNGSSFFTLAVOHRALVEGSAFPAVAVGSSCSWCDQAPPEILCYLOSFWTGSFLLANFDS 264
181 VNGSSFFTLAVOHRALVEGSAFPAVAVGSSCSWCDQAPPEILCYLOSFWTGSFLLANFDS 240

Qy	265	RSKCDANTLLGSIHTEDEPRAACDDSPQCSPRALANKKEVDSFRSIYTLNDGLSPSEA	334
Db	241	RSKCDANTLLGSIHTEDEPRAACDDSPQCSPRALANKKEVDSFRSIYTLNDGLSPSEA	300
Qy	325	VAVGRIPEDTYINGNFMPLCTLAABEQLYDALYOMDKGSLKVTDVSLDFPKALYSDAAT	3844
Db	301	VAVGRIPEDTYINGNFMPLCTLAABEQLYDALYOMDKGSLKVTDVSLDFPKALYSDAAT	360
Qy	385	GYSSSSSSSTYSIYDAVKTFADGPAFISIVETHAASNGSSEQYDSDGEGLSARDLTMSYA	4444
Db	361	GYSSSSSSSTYSIYDAVKTFADGPAFISIVETHAASNGSSEQYDSDGEGLSARDLTMSYA	420
Qy	445	ALLTANNRRNSVVPASMGETSASSVPGTCAATSAIGYSSVTVTSMPSIIVATGGTITTTAT	5040
Db	421	ALLTANNRRNSVVPASMGETSASSVPGTCAATSAIGYSSVTVTSMPSIIVATGGTITTTAT	480
Qy	505	PTGGSGVSTSKTTLTAAKSTSTTSS 530	
Db	481	PTGGSGVSTSKTTLTAAKSTSTTSS 506	

```

RESULT 9
US-09-199-290-34
: Sequence 34, Application US/09199290
: Patent No. 625084
: GENERAL INFORMATION:
: APPLICANT: Nielsen, Bjarne R.
: APPLICANT: Nielsen, Ruby
: APPLICANT: Lehbeck, Jan
: TITLE OF INVENTION: Thermostable Glucosylase
: FILE REFERENCE: 5279.200-US
: CURRENT APPLICATION NUMBER: US/09/199,290
: CURRENT FILING DATE: 1998-11-24
: EARLIER APPLICATION NUMBER: 1557/97
: EARLIER FILING DATE: 1997-12-30
: EARLIER APPLICATION NUMBER: 0925/98
: EARLIER FILING DATE: 1998-07-10
: EARLIER APPLICATION NUMBER: 60/070,746
: EARLIER FILING DATE: 1998-01-08
: EARLIER APPLICATION NUMBER: 60/094,344
: EARLIER FILING DATE: 1998-07-28
: EARLIER APPLICATION NUMBER: 08/979,673
: EARLIER FILING DATE: 1997-11-26
: EARLIER APPLICATION NUMBER: 09/107,657
: EARLIER FILING DATE: 1998-06-30
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 34
: LENGTH: 618
: TYPE: PRT
: ORGANISM: Talaromyces emersonii
US-09-199-290-34

```

[illegible]

```

OY      244  ILCYLOSQWTSFILIANT-DSGRSKDNANTLLGS:HTHPBEAACDSTQPCSPRALAH 3022
Db      248  VLCFQSGWTSQSYLVANFGSGRSKDVNSILGS:HTHPBAGCDSTQPCSPARALAH 3077

OY      303  KEVVDSPFSIYTLTNDGLSDSEAVANGRIPEDTYVNGNPMWFLCTLAAEOLYDALYOMXO 362
Db      308  KVVVDSPFSRIYAINSGIAEGSAVAAGRYPEDYVQGNPMYTLATAAAEOLYDALYOMKXI 3677

OY      363  GSLEVTDRSLPFPKALYSDAATGCTGSSSSSTSYSDVAWKTFADEGVSIYETHAASNGSM 4222
Db      368  GSIISTVDLSLFPFOIYPSAAVGTNNGSGTTINDIISANQYTGDTLSIVEXKTTPEBDGL 4277

OY      423  SEQYDKSDGEOLASDLITWSYALLITANNRRNSVVPASWGETSSASSVPGTCAATSAIGTY 482
Db      428  TEOEFERTIGTPLSASALITWSYASLLITASRRQSVVPASGESSASSASVPPVCASITSATGPY 4877

OY      483  SSVVYTSMPSIYATGCTTTLANPTGSGVTSRSTKTTATASKSTSTT 528
Db      488  STAITVWPS-----SSGSGSTTITSSAPCTTPTSVAYT 520

```

```

/ RESULT 10
/ US-09-821-616-34
/ Sequence 34, Application US/09821616
/ Patent No. 6620924
/ GENERAL INFORMATION:
/ APPLICANT: Nielsen, Bjarne R.
/ APPLICANT: Nielsen, Ruby
/ APPLICANT: Lehmebeck, Jan
/ TITLE OF INVENTION: Thermostable Glucosamylase
/ FILE REFERENCE: 5279.200-US
/ CURRENT APPLICATION NUMBER: US/09/821.616
/ PRIOR FILING DATE: 2001-03-29
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/199,290
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-24
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1557/97
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-30
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 0925/98
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-10
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,746
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-08
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/094,344
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-28
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/979,673
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-26
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/107,657
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-30
/ NUMBER OF SEQ ID NOS: 34
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 34
/ LENGTH: 618
/ TYPE: PRT

```

[illegible]

Db 188 QNDLSITTOYWNSTEDLWEEVSGSFFTTAVOHRALVBSGNALATRLNHTCSNCVSOAPQ 247
QY 244 ILCYLOSFWTGSFILANF-DSSRSGKDNANTLLGSIHTFPEAACDDSTFQPCSPRALANH 302
Db 248 VLCFLOSVMYTGSLVLANFGSGSGKDVNSILGSIHTFDPAGCCDDSTFQPCSPRALANH 307
QY 303 KEVYDSFRSITLINDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABOLYDALYOMDKO 362
Db 308 KVTLDPSRSITVAINSGIAEGSAVAVGRYPEDVYQGGGNPMFLATRAAAEQLYDAIYOMKKI 367
QY 363 GSIENVDSILDFPKALYSDAATCTSSSSSTSYDVAVTFADGFVSIYETHAANGSM 422
Db 368 GSIENVDSILDFPKALYSDAATCTSSSSSTSYDVAVTFADGFVSIYETHAANGSM 427
QY 423 SEQYDSDGEOLSGARLTMWSYALLTANNRRNSVVPASMGESASVPGTCATSAIGTY 482
Db 428 TEPFSITDGPPLSASALTMSYASLITRSAROSVVPASMGESSASVPAVCSATSGPY 487
QY 483 SSVTVTSWPSIVATGTTTATPTGSGSVTSTSKTATATASKTSTTT 528
Db 488 STATVWPS-----SGSGSSTTSSAPCTTPTSVAVT 520

RESULT 11
US-09-199-290-7
Sequence 7, Application US/09199290
Patent No. 6255084
GENERAL INFORMATION:
APPLICANT: Nielsen, Bjarne R.
APPLICANT: Nielsen, Rudy
APPLICANT: Lehmebeck, Jan
TITLE OF INVENTION: Thermostable Glucoamylase
FILE REFERENCE: 5279,200-US
CURRENT APPLICATION NUMBER: US/09/199,290
CURRENT FILING DATE: 1998-11-24
EARLIER APPLICATION NUMBER: 1557/97
EARLIER FILING DATE: 1997-12-30
EARLIER APPLICATION NUMBER: 0925/98
EARLIER FILING DATE: 1998-07-10
EARLIER APPLICATION NUMBER: 60/070,746
EARLIER FILING DATE: 1998-01-08
EARLIER APPLICATION NUMBER: 60/094,344
EARLIER FILING DATE: 1998-07-28
EARLIER APPLICATION NUMBER: 08/979,673
EARLIER FILING DATE: 1997-11-26
EARLIER APPLICATION NUMBER: 09/107,657
EARLIER FILING DATE: 1998-06-30
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 591
TYPE: PRT
ORGANISM: Talaromyces emersonii
US-09-199-290-7

Query Match 60.4%; Score 1674.5; DB 3; Length 591;
Best Local Similarity 62.3%; Pred. No. 1.5e-136;
Matches 314; Conservative 73; Mismatches 102; Indels 15; Gaps 3;

QY 26 TLDWSLNEATVARTAILNNGIGADGAMVSGADSGIVVSPSTNDPFIYTWTRDGLVTK 85
Db 4 SLDSPFLATEPPLALQGVLLNNGIGADGAMVSGADSGIVVSPSTNDPFIYTWTRDGLVTK 63
QY 86 TLVDLFRNDGTSLLSTIENYISAQAVQGISNPSGDLSSGAGIGBPKFVNDETAATYGSWG 145
Db 64 YLVDANRGNKMLEQITIQYISAQAVQGISNPSGDLSSGAGIGBPKFVNDETAATYGSWG 122
QY 146 RPDGDPALRATAMIGFGQWLLDNGYTSATDIWPLVPLVNDLSYVAQYNNQGYDLMEEV 205
Db 123 RPDGDPALRATAMIGFGQWLLDNGYTSATDIWPLVPLVNDLSYVAQYNNQGYDLMEEV 182
QY 206 NGSSFTIAVQHRALVEGSAFATAVSGSCGSCDQAPPEILCYLOSFWTGSFILANF-DSS 264

Db 183 EGSSFTTAVQHRALVEGSAFATAVSGSCGSCDQAPPEILCYLOSFWTGSFILANFGSG 242
QY 245 RSGKDNANTLLGSIHTFPEAACDDSTFQPCSPRALANHEVDSFRSITLINDGLSDSEA 324
Db 243 RSGKDNANTLLGSIHTFPEAACDDSTFQPCSPRALANHEVDSFRSITLINDGLSDSEA 302
QY 325 VAVGRYPEDTYNGNPMFLCTLAABOLYDALYOMDKOSLEVTDSLDFPKALYSDAAT 364
Db 303 VAVGRYPEDVYQGGGNPMFLATRAAAEQLYDAIYOMKKISITIDVSLPFQDIYPSAAV 362
QY 365 GTSYSSSTSYDVAVTFADGFVSIYETHAANGSMGSKGQKSGEOLSGARLTMWSYA 444
Db 363 GTSYSSSTSYDVAVTFADGFVSIYETHAANGSMGSKGQKSGEOLSGARLTMWSYA 422
QY 445 ALLTANNRRNSVVPASMGESASVPGTCATSAIGTYSSVTVTSWPSIVATGTTTAT 504
Db 423 SLITTSAROSVVPASMGESSASVPAVCSATSGPYSTATVWPS----- 470
QY 505 PTGSGSVTSTSKTATATASKTSTTT 528
Db 471 -SGSGSSTTSSAPCTTPTSVAVT 493

RESULT 12
US-09-821-616-7
Sequence 7, Application US/09821616
Patent No. 6620924
GENERAL INFORMATION:
APPLICANT: Nielsen, Bjarne R.
APPLICANT: Nielsen, Rudy
APPLICANT: Lehmebeck, Jan
TITLE OF INVENTION: Thermostable Glucoamylase
FILE REFERENCE: 5279,200-US
CURRENT APPLICATION NUMBER: US/09/821,616
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/199,290
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-24
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1557/97
PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-30
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 0925/98
PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-10
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,746
PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-08
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/094,344
PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-28
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/979,673
PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/107,657
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-30
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 591
TYPE: PRT
ORGANISM: Talaromyces emersonii
US-09-821-616-7

Query Match 60.4%; Score 1674.5; DB 4; Length 591;
Best Local Similarity 62.3%; Pred. No. 1.5e-136;
Matches 314; Conservative 73; Mismatches 102; Indels 15; Gaps 3;

QY 26 TLDWSLNEATVARTAILNNGIGADGAMVSGADSGIVVSPSTNDPFIYTWTRDGLVTK 85
Db 4 SLDSPFLATEPPLALQGVLLNNGIGADGAMVSGADSGIVVSPSTNDPFIYTWTRDGLVTK 63
QY 86 TLVDLFRNDGTSLLSTIENYISAQAVQGISNPSGDLSSGAGIGBPKFVNDETAATYGSWG 145
Db 64 YLVDANRGNKMLEQITIQYISAQAVQGISNPSGDLSSGAGIGBPKFVNDETAATYGSWG 122
QY 146 RPDGDPALRATAMIGFGQWLLDNGYTSATDIWPLVPLVNDLSYVAQYNNQGYDLMEEV 205
Db 123 RPDGDPALRATAMIGFGQWLLDNGYTSATDIWPLVPLVNDLSYVAQYNNQGYDLMEEV 182

Qy	206	NSSPFTLIVORHALVEGSAFATPAVAGSSCMWCSOAPILICYLOSFWMGSTLIANF-DSS	264
Dd	183	EGBSPFTTIAVOHRAIVEGNALATRLNHTCNSCVSAPVOLFLCOSYMTGSVI LANFGSG	242
Qy	265	RSGKDNLTLLGISIHFPDEPAACDDBSTFOPCSPRALANKEVDSFRSIYTLNDGLSDSEA	324
Dd	243	RSCKDVNSTLIGSIHFHPDPAGCGDDBSTFOPCSARLANHKVTDSPRISYI LANSGI A GSA	302
Qy	325	VAVGRYPEDTYNGNEMFLCTIAAAEOLYDALYOMDKOSLEVTDIVSLDFPKALYSDAAT	384
Dd	303	VAAGRYPEDVYQGCGNPMWLATAAAAEOLYDALYOKKIGISIITDVSLPFPODIYPSAAV	362
Qy	385	GITYSSSSSYTSYLVDAVKTFADGPFSIYEYTHAASNGSMSSEODXSDERQSARPLUTSYA	444
Dd	363	GIYNSSGTTFNDIISAAYOTYGDGYLSIEKYPPSPSGSLTEOFSTRSDGPIASALATWSYA	422
Qy	445	ALLTANNRRNSVPASNMGEITSASSVPGCAATSALICTYSSVTYVMSWPBIVATGGTTTTAT	504
Dd	423	SILLTAARRSRVVPAPAWGESSASSSVLAVCASATSADGPSTRTNRVWP-----	470
Qy	505	PFGSGSVTPSKTTATASKTSTTT	528
Dd	471	-SGSGSSTTSSACPCTPTPSVAVT	493

```

RESULT 13
US-09-704-449-2
; Sequence 2, Application US/09704449
; Patent No. 6309872
; GENERAL INFORMATION:
; APPLICANT: Rey, Michael W.
; APPLICANT: Golihtly, Elizabeth J.
; TITLE OF INVENTION: Polypeptides Having Glucosylase
; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 10101.000-US
; CURRENT APPLICATION NUMBER: US/09/704,449
; CURRENT FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Thielavia terrestris
US-09-704-449-2

```

Query Match	58.3%;	Score 1615.5;	DB 4;	Length 630;
Best Local Similarity	58.9%;	Pred. No. 2.2e-131;		
Matches 315;	Conservative 83;	Mismatches 112;	Indels 25;	Gaps 7

Qy	6	LLALSGVCTGGLANVLSKRA-----TLDSTMSNATVARKTILINNIQADAWVSGA	56
Db	6	LIGLHLLPALALGHPEASRYRREGEVYVKSVDSPFATESPALASMLLCNIGSTGCHASGV	65
Qy	57	DSGLVVASPESTNDPQFYFTWTRDSGLVLTVDLRNG-DTSLSTIENYISQAQIVQGI	115
Db	66	ASGLVVASPKTKPTMDWYTWTRDSALTFPCQVDPTNSYDSLQEIOMNYVVAAPHIQGV	129
Qy	116	SNPSGDISSAGLGEEKFENVDETAAYGSGMRPQRPQALRAITAMIGFCQMLLDNGYSTA	175
Db	126	SNPSGSLSDSGSGEGEKFENVDMQSPFGAMGRQRRGPPALRALALAYSXWILNSGYSTA	189
Qy	176	TDIWPPLVLRNDLSYVAQYNQGTGYDLMEEVNGSSFFTLAVQHRALVEGSAFATVSSCS	235
Db	186	SSLIWPPVTKIDLAYVQO--NNTGPFDMEEVSSSFFTYANQHRALVEGSAFATVSLGTS	243
Qy	236	WCDSQAPPELICYSQSFMTGS--FTLANPSSRSRGDANTLIGSIHTPPEAACDSTFOP	293
Db	244	ACSAVAPQILCFIQSFMTSPSSGYILAN-----STADANTLIGSIHTPPEAACDAPFP	299
Qy	294	CSPPALANKEVVDFFRSITLYTLNDGLSDSEAVANGRYPEDTYINGNPMPFLCTLAAEOLY	355
Db	300	CSBRALANHKVVDARFSYISINSIGAESAAVAKRPEPDSYFGGNPMPVINTLTLAAQOLY	355

```

OY      354  AALVQWMDKQSGLEIETVDVLDLDFEKKALYSPAATSTYSSSSSTYSIIDAVATGDFPJSIYE 413
Db      360  DALIYWKQSGSTYTSTSLAFFPKDSSSTTTPETYSSTSTYTTLLNAISAYADGKNYIA 419

OY      414  THAASNGSMSEQYDKSDGEQLSARDLTWWSYALLTANNRRNSVVPASWGETSASSVPGTC 473
Db      420  QYAQNGSLSEQFSKTNGEPLSAYLLTWSYALFLTPAARRAGVVPWSGMAASNSVPAQC 479

OY      474  AATSAIGYSSVTVWSWPSIATVATGTTTTTAAFPDGSQSVTSTSKTTATASKTSTTT 528
Db      480  SATSVSGYSTATATSPFP-----SQTPASSTSMQS--SPASTTATATACSTPT 527

```

RESULT 14
 US-08-596-300A-7
 Sequence 7, Application US/08596300A
 Patent No. 5834191
 GENERAL INFORMATION:
 APPLICANT:
 TITLE OF INVENTION: Production of Heterologous Peptides
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
 CURRENT APPLICATION DATA:

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Query Match	56.9%	Score 1577;	DB 2;	Length 626;
Best Local Similarity	57.6%	Pred. No. 4.8e-128;		
Matches 306;	Conservative 85;	Mismatches 130;	Indels 10;	Gaps 6;

QY 2 SPRLSLTSGVLCTGTLANVISRATLIDSWLSNEATFARPAIILNNICADGAWMSGAGSIV 61
 Db 13 AFQAVLGLPDPPLAHEKRHSIIKR-SVDSYIIQETPIPAQKULCNICASGCRASGAASGV 71
 QY 62 VASPSLDPDPVYFTWRDPSGLVLTLDVDFRNG-DTSLSTIENYISAQAVOIGINPSG 120
 Db 72 VASPSKSPDYIYTTTRDAALVTKLIVDEFMTDYNITTLQNTICAYAAQAQKQGVENPSG 131
 QY 121 DLSGAGLGEPEKFNVDETAYTGSWGRPORHGALRATAMIGFGOMLLDNGYSTATDIW 180
 Db 132 SLISNAGLGEPEKFNVDLOQFTGAMKRPQGDGPRPLRALIIGVKMLVSNAGYATASLIW 191
 QY 181 PLVRNDLSVAQYMNQGTGYDLMEVNGSSFTIAYQHRALVEGSAFATAVGSSCSMCDQ 240

Db 192 PIVKNDLAVTAQVMNNTGFDLWBEVNSSFFLIASHRALVBSAFKSVSSCSACDAl 251
 Qy 241 APFIILCYLOSFWTGS-FILANFDSRSRGKADANTLIGSIHTTDPPEACDSTPQCSRAL 299
 Db 252 APQILCFQOSFWSNSGYIISNFVNVRSGDINSVLTSHNFDPAAGCQVNFQCSDRAL 311
 Qy 300 ANHKEVVDSPRSIYLTINDGLSDSEAVAVGRYEDTYNGNPMFLCTTAAAEQLVDALYOM 359
 Db 312 ANHKVAVDMSR-FWGNVNSRGTAKGAANGRYAEDYYINGNPMYLTATLAAEQLDAYVM 370
 Qy 360 DKQGSLEVTDVSLDFPKALVSDAATGYSSSSSIYSSIVDAVKTAFDQFVSIYETHAASN 419
 Db 371 KRQGSILVTSTSLAFKDLVPSVSGTYSSSSSTYALINAVTYADQFVDIYQYRPD 430
 Qy 420 GMSSEQYKXSDGEQLSARDULTVSYAALLTANNRRNSVVPASMGETSSASVGTCAATSAI 479
 Db 431 GSIAEOPFDXSGPILSATHLTWISYASFLSAAARRAGIIVPSPWGAASANSLPESGSASTVA 490
 Qy 480 GYTSVTVTWSPSIVATGGTTT-ATPGSGS-----YTSRSKTTATASKT 524
 Db 491 GSAATATATSPANLTPASTYVTPPQIGCAADHEVLVTNEKTTYSIGQT 541

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

Query March      56.9%  Score 1577:  DB 2      Length 626;
Best Local Similarity  57.6%  Pred. No. 4.8e+28;
Matches 306;  Conservative  85;  Mismatches 130;  Indels  10;  Gaps  6;

OY      2  SFRSLLLSGLVCGIANVISRATLDLSMSLEATVARTAILNNGADGAWNSGADSGIV 61
:|||||  :  : ||| :||| :  : :  : |||||  ||||| :||

```

[illegible]

```

RESULT 16
US-09-534-407-3
; Sequence 3, Application US/09534407
; Patent No. 6361973
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Kimberly Brown
; TITLE OF INVENTION: Promoters For Expressing Genes In A
; TITLE OF INVENTION: Fungal Cell
; FILE REFERENCE: 5611.200-US
; CURRENT APPLICATION NUMBER: US/09/534.407
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/274.449
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 581
; TYPE: PR1
; ORGANISM: Fusarium
; US-09-534-407-3

```

[illegible]

```
QY      188 RNDLSVAAYQNMONTGVDLMEENSGSFFFLIANQHIALVEGAFAFATVAGSCSMWDSQAPE 243
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      185 EKOLATTYYKMNNTGYDLMEEVNGSSFLLSHSHALVEGAAIAKRGSCTPDVTNAIPR 244

QY      244 ILCYLIOFSFTMGSEFILNPF--SSRSKGDAANTLLGISIHFPDEAACDDSTFOPCSPRALAN 3010
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      245 VLCLFIQTFTMTGGVSDSNINVKGRGLDVNSTLSSIHFPDENSCKTSTPOCSPRALAN 3040

QY      302 HKEVDSFRSIYTLANDGLSDSEAVAVGRYPEDTYNGNPWF.LCTLAABEQLYDAL.YOMDK 3610
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      305 HKAVVDSFRSIYGVNNKNRGOKAAAVAGRYSEDVVYIDGNPWLTATILAAAEQLYAIVYOMDK 3640

QY      362 QGSLAEVTDVSLDFPKALLYSDAATGTYSSSSSSYSSI.VDAVKT.PADGFVS.IYETHAASNGS 4210
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      365 LGAVTDVDDVSLSEFKDIVPVRKSJGYAKTKTKYEII.KAAATYADVGEFAVAVQTTPPDGS 4240

QY      422 MSBQVYKSDQSEQLSADLT.WMSYAFALLITANNRNNSVPVPSKWETNASASSPGICAAITS.IGT 4810
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      425 LAEOFPDKSTGAPKSAAH.LTWMSYAAFAFATTERDRDIGISPMWESSANKRVPAVCOAPA.CDT 4840

QY      482 -----YSSVTVTWSPI.SI.VANGGTTTA-----IPTSG 509
        :|:||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      485 TTFESVANVOVSSDQKVTVVGSVTELSMKSPPDGIALTPSSSG 527
```

```

RESULT 17
US-09-999-201B-4
: Sequence 4, Application US/09999201B
: Patent No. 6518044
: GENERAL INFORMATION:
: APPLICANT: Berka, Randy
: APPLICANT: Rey, Michael
: APPLICANT: Brown, Kimberly
: TITLE OF INVENTION: Promoters For Expressing Genes In A
: TITLE OF INVENTION: Fungal Cell
: FILE REFERENCE: 5611.210-US
: CURRENT APPLICATION NUMBER: US/09/999,201B
: CURRENT FILING DATE: 2001-10-30
: PRIOR APPLICATION NUMBER: 09/534,407
: PRIOR FILING DATE: 2000-03-22
: PRIOR APPLICATION NUMBER: 09/274,449
: PRIOR FILING DATE: 1999-03-22
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 581
:
: TYPE: PRF
: ORGANISM: Fusarium
: US-09-999-201B-4

```

Query Match	51.4%;	Score 1425.5;	DB 4;	Length 581;
Best Local Similarity	52.2%;	Pred. No. 5.8e-115;		
Matches 273;	Conservative 81;	Mismatches 146;	Indels 23;	Gaps 6

[illegible]

```

Db      245 VLFLQTLPTMGTVGVVSNINVKDGRGRGLDVNSILISIHFDENSKCTDSTFQPCSPALAN 304
Qy      302 HKRVVDSFSSITLTLDGLSDSEAVAVERYPETTYINGNPWFLCTLLAAEQLYDALYWDK 361
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      305 HKVVDVPSRSITGVNNKRGQRAAVVGRISBVDVYDGNPWLATLLAAEQLYAIVVWDK 364
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      362 QGSELEVTVDSLDFPKALYSDAATGYFSSSSSTYSISIVAVTYPADGFASIVETHAASNGS 421
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      365 LGAVYVDVDSLSPFKDVIYPKSKGYAKTKYKELIAATYADGFVAVVGQITTPRDS 424
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      422 MSEQVYKSDGEBQLSARDLTWSTYALLLTANNRNSVVPASWGETPSSVPGTCAATSIGT 481
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      425 LAEQPKFSKGPAPKSAVHLMTSYAAAEVAATTERDGLISPMWESSANKVPAPVCAAPACDT 484
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      482 -----YSSVTVTSMPEIVATGGTTT-----PTGSG 509
         : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      485 TTFPSVKNVQVSSDQKTVYVGSVTELSNMSPPDGIALTTPSSSG 527

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	47.7%;	Score 1321;	DB 1;	Length 616;
Best Local Similarity	48.6%;	Pred. No. 7.3e-106;		
Matches 257;	Conservative 88;	Mismatches 150;	Indels 34;	Gaps 6;


```

QY 14 CTGLANVIS-----KRATLDSWLSNEATVARTAILNINIGADGAWVSGADSGIYVA 63
DB 8 CAGALSILCSLAIAAPTELKARDLSFISERAIALQALNINIGPDGSAVPAGAFVYA 67
QY 64 SPSKAPDVFYTWTRDSGLVLTVDLFRNGDLSLSTIENTYISAGAIYOGISNPGSD-L 122
DB 68 SPSKAPDVFYTWTRDSGLVLTVDLFRNGDLSLSTIENTYISAGAIYOGISNPGSD-L 127
QY 123 SSGAGIGEPKFNVDERTAYTGSWGRPORDPALRATMIGGQWLDNGYSTATDVIWPL 182
DB 128 PDGVGIGEPKFNVDERTAYTGSWGRPORDPALRATMIGGQWLDNGYSTATDVIWPL 187
QY 183 VRNDLSYVAQYNNQTYDLMEEVNGSSFTTIAVQHRALVEGSAFATAVSSCSWCDSOAP 242
DB 188 IANDLSYVQYNNQSFDMEEETVYASSFTTIOQHRAVLEGAQALMDLVCTGCC-QAP 246
QY 243 EILCYLOSFWTGSFILANF--DSSRSGKDANTLGSITHTDPEAACDSTFQPCSPRALA 300
DB 247 EYLCLFQSFNGKVIYISNINNNNGRTGLDGNISLGAISTFDIDAYCDSPFLQCHQSOLA 306
QY 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNWPFLCTLAABQLYDALYOND 360
DB 307 NFKVLDTFRNLTYTINAGIPEGGVAVGRYAEVDYNGNWPFLCTLAABQLYDALYOND 366
QY 361 KQGSLEVDVSLDFEPALYSDAATGYSS--SSSTYSIVDAVKTPEADGFSIVETHAAS 418
DB 367 ARHVLTVDBETSLAFEDYIPEVTREYKSGNNSPPAQIMDAVTAADSVVAIAEKTIIPS 426
QY 419 NSGMEQYKSDGEOQSARDLTWSYALLTANNRNSVVPASGGETSASSVPGTCAATSA 478
DB 427 NSLSQFNRDGTPLSAIDLWTMSYAAFITMSQRRAGQYPSGWSGNALPPTTCSASST 486
QY 479 IGTYSVTVTSMPSIVATGTTTATPTGSGSVTSTKTATASKTST 527
DB 487 PGTY-----TPATAAGAPNVTSSCQVSIITFINATT 517

```

RESULT 19 US-08-385-370-4

```

; Sequence 4, Application US/08385370
; Patent No. 5665585
; GENERAL INFORMATION:
; APPLICANT: Joutskeli, Tuula
; APPLICANT: Joutskeli, Vesa
; APPLICANT: Torkkeli, Helena
; APPLICANT: Vainio, Arja
; APPLICANT: Fagerstrom, Richard
; APPLICANT: Aho, Sirpa
; APPLICANT: Korhola, Matti
; APPLICANT: Nevalainen, Helena
; TITLE OF INVENTION: Production of Glucoamylase P with High
; TITLE OF INVENTION: Limit Dextrinase Activity in Trichoderma
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/385,370
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/104,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Cimbalá, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1050.0270004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 616 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-385-370-4

```

```

Query Match 47.7%; Score 1321; DB 1; Length 616;
Best Local Similarity 48.6%; Pred. No. 7,3e-106;
Matches 257; Conservative 88; Mismatches 150; Indels 34; Gaps 6;

```

```

QY 14 CTGLANVIS-----KRATLDSWLSNEATVARTAILNINIGADGAWVSGADSGIYVA 63
DB 8 CAGALSILCSLAIAAPTELKARDLSFISERAIALQALNINIGPDGSAVPAGAFVYA 67
QY 64 SPSKAPDVFYTWTRDSGLVLTVDLFRNGDLSLSTIENTYISAGAIYOGISNPGSD-L 122
DB 68 SPSKAPDVFYTWTRDSGLVLTVDLFRNGDLSLSTIENTYISAGAIYOGISNPGSD-L 127
QY 123 SSGAGIGEPKFNVDERTAYTGSWGRPORDPALRATMIGGQWLDNGYSTATDVIWPL 182
DB 128 PDGVGIGEPKFNVDERTAYTGSWGRPORDPALRATMIGGQWLDNGYSTATDVIWPL 187
QY 183 VRNDLSYVAQYNNQTYDLMEEVNGSSFTTIAVQHRALVEGSAFATAVSSCSWCDSOAP 242
DB 188 IANDLSYVQYNNQSFDMEEETVYASSFTTIOQHRAVLEGAQALMDLVCTGCC-QAP 246
QY 243 EILCYLOSFWTGSFILANF--DSSRSGKDANTLGSITHTDPEAACDSTFQPCSPRALA 300
DB 247 EYLCLFQSFNGKVIYISNINNNNGRTGLDGNISLGAISTFDIDAYCDSPFLQCHQSOLA 306
QY 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNWPFLCTLAABQLYDALYOND 360
DB 307 NFKVLDTFRNLTYTINAGIPEGGVAVGRYAEVDYNGNWPFLCTLAABQLYDALYOND 366
QY 361 KQGSLEVDVSLDFEPALYSDAATGYSS--SSSTYSIVDAVKTPEADGFSIVETHAAS 418
DB 367 ARHVLTVDBETSLAFEDYIPEVTREYKSGNNSPPAQIMDAVTAADSVVAIAEKTIIPS 426
QY 419 NSGMEQYKSDGEOQSARDLTWSYALLTANNRNSVVPASGGETSASSVPGTCAATSA 478
DB 427 NSLSQFNRDGTPLSAIDLWTMSYAAFITMSQRRAGQYPSGWSGNALPPTTCSASST 486
QY 479 IGTYSVTVTSMPSIVATGTTTATPTGSGSVTSTKTATASKTST 527
DB 487 PGTY-----TPATAAGAPNVTSSCQVSIITFINATT 517

```

RESULT 20

```

; US-08-270-076A-11
; Sequence 11, Application US/08270076A
; Patent No. 5667986
; GENERAL INFORMATION:
; APPLICANT: Sleep, Darrell
; APPLICANT: Goodey, Andrew R
; APPLICANT: Vakeria, Diana
; TITLE OF INVENTION: Yeast Promoter
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The BOC Group, Inc.
; STREET: 100 Mountain Avenue, Murray Hill
; CITY: New Providence
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07974

```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/270,076A
FILING DATE: 01-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8923521.2
FILING DATE: 18-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/597,687
FILING DATE: 16-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/925,286
FILING DATE: 04-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Swope, R. Hain
REGISTRATION NUMBER: 24864
REFERENCE/DOCKET NUMBER: 92H834-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908/771-6292
TELEFAX: 908/771-6159
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 806 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-270-076A-11

Query Match      21.3%; Score 589; DB 1; Length 806;
Best Local Similarity 30.6%; Pred. No. 3,3e-42;
Matches 154; Conservative 90; Mismatches 200; Indels 60; Gaps 15;
```

```
1 MSFRLSLALSGLVCTGLANVSKRATLDSWLSNEATVARTAILNIGADGAWVSGADSGI 60
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
313 VQLADVIVAMNGTVYVD-SNGAMDSSALBEMLQROKVVSIERIFENIGSAVYPS-ILPGV 370
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 VVASPSTNDNRYFTYTRDLSGLVKTIVDLFRNDTSLSTLSTLSTLSTLSTLSTLSTLST 120
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
371 VIAPSPQTHDPYFQWIRDSALTTNSIVS--HSADPA-IETLQYLVNVSFHLQRTNN--- 424
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 DLSSGAG-----LGEPEVNDETAYTGSWGRFQPDGPALRAVAMIGFGWLDNG--- 170
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425 TLGAGIGYTDVYALGPPKMNVDNTAFTPEWGRFQNDGPALRSIAIILKIIDYIKQSGTDL 484
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
171 -----YTSATDIWPLVRNDLSVAAQYMNQGYDLMEVNGSFFTIYVQHRALVVGSA 225
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
485 GAKYPFQSTA-DIPDDIVRWYLRFIIDHMNSSGFDLMEVNGHFFTLVQLSAVDRTLS 543
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
226 FATVAGSSCSWCD---SQAPEILCYL-----QSFWTGSFILANF-----DSRSQKDA 270
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
544 YFNASERSSPVEELRQTRRDISKFLVDPANGFINGKY---NYIVETPMIADTLRSGLDI 600
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
271 NTLLGSIHTPEPEAACDDSTFQPCSPRALANHKKEVDSFRSITYTLNDGLSDSEAVAVGRY 330
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
601 STLIANTVHDAPSA-SHLPFDINDPAVLNTLHLMHMSISYIINDSSKNATGIALGRY 659
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
331 PEDTY-----YNGNPWFLCTIAAEOLYDALYQMDKQSLSEVTVDISLDFKALYSDATG 385
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
660 PEDYDYGVGVEGHPWLATCAASTLYQLIYRIHQHDLVVMNNDCSNAFWSELVFS 719
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
386 TYSS-----SSSTYSIVDAVKTPADGFVSIVETHAASNGSMGQYKXSGEQ 433
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
720 NLTLTGDEGYLLIEFTYPAFNQTIQKIFQLADSLFLVKAHVGTDGDELSEQFNKYYTGM 779
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
434 LSARDLTMSYAAALLTANRRNSV 457
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
780 QGAQHLTWSTYSFMDAYQIROEVL 803
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 21
US-09-199-290-4
Sequence 4, Application US/09199290
Patent No. 6255084
GENERAL INFORMATION:
APPLICANT: Nielsen, Bjarne R.
APPLICANT: Nielsen, Rudy
TITLE OF INVENTION: Thermostable Glucoamylase
FILE REFERENCE: 5279,200-US
CURRENT APPLICATION NUMBER: US/09/199,290
CURRENT FILING DATE: 1998-11-24
EARLIER APPLICATION NUMBER: 1557/97
EARLIER FILING DATE: 1997-12-30
EARLIER APPLICATION NUMBER: 0925/98
EARLIER FILING DATE: 1998-07-10
EARLIER APPLICATION NUMBER: 60/070,746
EARLIER FILING DATE: 1998-01-08
EARLIER APPLICATION NUMBER: 60/094,344
EARLIER FILING DATE: 1998-07-28
EARLIER APPLICATION NUMBER: 08/979,673
EARLIER FILING DATE: 1997-11-26
EARLIER APPLICATION NUMBER: 09/107,657
EARLIER FILING DATE: 1998-06-30
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 35
TYPE: PRT
ORGANISM: Talaromyces emersonii
US-09-199-290-4

Query Match      5.8%; Score 160; DB 3; Length 35;
Best Local Similarity 80.0%; Pred. No. 3,4e-07;
Matches 28; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
269 DANVLGSIHTPEPEAACDDSTFQPCSPRALANHK 303
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 DVNSILGSIHTFPAGCDDSTFQPCSPRALANHK 35
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 22
US-09-821-616-4
Sequence 4, Application US/09821616
Patent No. 6620924
GENERAL INFORMATION:
APPLICANT: Nielsen, Bjarne R.
APPLICANT: Nielsen, Rudy
TITLE OF INVENTION: Thermostable Glucoamylase
FILE REFERENCE: 5279,200-US
CURRENT APPLICATION NUMBER: US/09/821,616
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/199,290
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-24
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1557/97
PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-30
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 0925/98
PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-10
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,746
PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-08
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/094,344
PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-28
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/979,673
PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/107,657
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-30
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 35
TYPE: PRT
```

ORGANISM: Talaromyces emersonii
US-09-821-616-4

Query Match 5.8%; Score 160; DB 4; Length 35;
Best Local Similarity 80.0%; Pred. No. 3.4e-07;
Matches 28; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 269 DANTLLGSIHTPDPEACDSTFPCSPRALANHK 303
Db 1 DVNSILGSIHTPDAGCDSTQPCSPRALANHK 35

RESULT 23

US-09-134-001C-4463
Sequence 4463, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

CURRENT FILING DATE: US/09/134,001C

PRIOR FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4463

LENGTH: 2137

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4463

Query Match 5.6%; Score 156.5; DB 4; Length 2137;
Best Local Similarity 21.7%; Pred. No. 0.00048;
Matches 126; Conservative 92; Mismatches 284; Indels 79; Gaps 17;

Qy 2 SFRSLALGLVCTGLANYSKRATLDSML-----SNATYARTRAILNIGADGAWVS 54
Db 1368 SEASATSLSGSTSLSDSTS-TSTSDASSTGVSVDNSASTSLSGSLSTSVSDSTS 1426

Qy 55 GADSGIVASPTDNDYFYTWTRDGLVKTLDVDFRNGDTSLST---INYSIAQA 110
Db 1427 TSSS-----ASATSESDSERASTSLSGSTSLSDSTSSTSDASTSVSSNSTSTS 1482

Qy 111 IVQGISNPGDLSGAGLGPKNVDETAYTSGMGRPORDGAPALRATAMIGCOWLIDNG 170
Db 1483 ISRSLSSTSVSDSTS-----TSTSDAST-STSVSDSASTSVSSSVSTS---DSE 1529

Qy 171 YNSTADIWPLVRNDLSYVAQWNOTGYDLMEEVNGSFFTAVQHRALVEGSAFATAY 230
Db 1530 STSTSTS-----DSASTSVSSNSTSTSL---SGSTSVSDSTSSTSDASTS 1580

Qy 231 GSSCSMCDGAPFLLCYLGFMTGSIPLANFDSRSGKQANTLLGSIHTPDPAACDDST 290
Db 1581 ESDSDASTSSSSSVSTSVSDSTSSTSDASTSVSDNSASTSL-SESTSTSLSDST 1639

Qy 291 POPCSPRALANHKVVDSPRSIYTLNDGLSD--SEAVAVGRYPEDTYNGNPMFLCTLA 348
Db 1640 SMSTSDASTSTSE-SDSDASTSLSDSTSVSESTSTSVSAANSTSTSLSDSR 1698

Qy 349 AEGULYALYQMDQGLSEVTVDSLDFPKALYSDAATGYSSGST-----393
Db 1699 TSLSDSTSTSTSESGSTSVSDSDASTSLSDSTSTSTSTSTSDASTSVSSVS 1758

Qy 394 ---YSGIVDAVKTFADGPFVIVETHAASNG-----SMSEQDKSDGQLSARDL 439
Db 1759 NRASSTSLSDSTSVSDSTSVSDSTSVSDSTSVSDSTSVSDSTSVSDSTSVSDST 1816

Qy 440 TWSYALALTNANRRNSVVPASWETSASVPGTCAATSA-IGTVSSVTYVWPSIVATGG 498
Db 1817 TSDSASTSTSESDSSTSTSLSDSTSVSDSTSVSDSTSVSDSTSVSDSTSVSDSTSLSGS 1876

Qy 499 TTT-----TATPTGSGSVTSTKTTATASKSTTTTSGMSL 534
Db 1877 TSTSVSDSTSI 1917

RESULT 24

US-09-463-712C-10
Sequence 10, Application US/09463712C
Patent No. 6558937

GENERAL INFORMATION:

APPLICANT: DSM, N.V.

APPLICANT: Gielkens, Marcus

APPLICANT: Vessier, Jacob

APPLICANT: De Graeff, Jaendert

TITLE OF INVENTION: CELLULOSE DEGRADING ENZYMES OF

FILE REFERENCE: 24615-20135.00

CURRENT FILING DATE: US/09/463,712C

CURRENT FILING DATE: 2000-04-04

PRIOR APPLICATION NUMBER: PCT/EP98/05047

PRIOR FILING DATE: 1998-07-31

NUMBER OF SEQ ID NOS: 14

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10

LENGTH: 536

TYPE: PRT

ORGANISM: Aspergillus niger

US-09-463-712C-10

Query Match 5.1%; Score 142; DB 4; Length 536;
Best Local Similarity 22.6%; Pred. No. 0.00096;
Matches 89; Conservative 48; Mismatches 136; Indels 120; Gaps 20;

Qy 205 VNGSFFTTAVOHRALVEGSAFATAVAGSSCSWCDGAPFLLCYLGFMTGSIPLANFDS- 263
Db 161 LMGALYFVAMDADGCTSEYSGNKAQYGTGYCDSCCPDLKFTNG-----EANCDCW 213

Qy 264 SRSQKANTLLG-----SIHTPDPAACDDSTFPCSPRALANHKVVDSPRSIYTLN 316
Db 214 EPRSNNVNVGVDPDHGSCAEMDVMEANSISNAFTAHPC-----DSVSQTMCDG 261

Qy 317 DGLSDSEAVAVGRY-----PEPTYN---GN-----PMFLCTLAARELY 353
Db 262 DSCGTYGYSAGRYSTCPDPCDVPYRLGNTDFPGRLGYDTNPFVYVTOFTD--- 318

Qy 354 DALYQMDKQGLEVTVDVSLDFPKALYSD-----AATG--TYSS--SSSTYSIVDAVKTPA 405
Db 319 -----DGTSSGTLTEI-----KRLVQNGEVIANCASTSVNGSSITSAPCESKTLF 367

Qy 406 DGFVSIIVETHAASNGSMSEQYDKSDGEQLS-----ARDLTWYALALLTNANRRNSVVPAS 460
Db 368 -GDENVFDKGLGLEG-MGEAMAKGVLVLMDVDYAADMLM-----LDSQ 410

Qy 461 WGETSASVP-----GTCATSAIG-----TYSSVTYSWPSIVANGGTTTAT 504
Db 411 YPNSSASTPGVARGTCSDSGVPATVEASPNAYTYSNT-----KCPGISSTSSGS 464

Qy 505 PTGSGSVTS-----TSKTTATASKTSTTTTSGMS 533
Db 465 SSGSGSSSSSSSTTKATSTTLKTTSTTSSGS 497

RESULT 25

US-08-614-377A-7
Sequence 7, Application US/08614377A
Patent No. 5976864

GENERAL INFORMATION:

APPLICANT: Smith, John

APPLICANT: Bingle, Wade H.

APPLICANT: No. 5976864ellini, John F.

TITLE OF INVENTION: EXPRESSION AND SECRETION OF

TITLE OF INVENTION: HETEROLOGOUS

```

; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson PC
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/614,377A
; FILING DATE: 12-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; APPLICATION NUMBER: US 07/895,367
; FILING DATE: 09-JUNE-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsao, Y. Rocky
; REGISTRATION NUMBER: 34053
; REFERENCE/DOCKET NUMBER: 08106/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1026 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-614-377A-7

Query Match      4.9%; Score 136; DB 2; Length 1026;
Best Local Similarity 23.2%; Pred. No. 0.009;
Matches 123; Conservative 73; Mismatches 225; Indels 110; Gaps 23;

QY 26 TLDSWLSNEATVATATILNNGADGAMVSGADSGIVASPTDNPDPFYTWTRDGLVVK 85
DB 28 TLDAAY-----ATQOTGGLSDAALTLTKLVNSTTAAVATQTYQ-----PFTGVADPSAAGLD 79
QY 86 TLVDLFRNGDTSLSLTIENTYISAQAIYQGISNPSGDLSSGAGLGEPRFNVDETAATYGSWG 145
DB 80 FLVDSTTN--TNDLN--DAYSKFAQENRFINFSINLATAGAGATAF--AAAYT----- 128
QY 146 RPQRDGPALATAMIGGQWLLDNGYTSATDVIWPLVRNDLSVYAOYMNQGYD-LMEE 204
DB 129 -----GVSYAQVATAYDKIIIGNAVATAAGVDVA-----AAVAFLSRQANIDYLTAF 175
QY 205 VNGSSEFTIAVQHRALVEGSAFATAVSGSCSCWCDGSAPELLCYLOSFWTGSFILANFDS 264
DB 176 VRANTPPTAADDIDLAVKALIGITILNAA-----TVSGIGGYATATAAMIN-DLS 224
QY 265 RSGKDANTLLG-SIHFPDPEAACDDSTFPQCSPPRALNHKEVDFSRSIYTLNDGLDSE 323
DB 225 DGLALSTNMAAGVNLFTAYPSSGVSGSTL-----SLTGTDTTLT--- 262
QY 324 AVAVAGRPEDTYNGNFWFLCTLAABQLYDALYQMDKQSLLEVTVSLDFKALYSDA- 382
DB 263 -----GTANNDTFVAGEVAGATTLVGGTLLSGG-----AGTIDV-LMWVQAAVATL 307
QY 383 ATGYSSSSSTYSIVDAVKTF--ADGFVSIIVETHAASNGMSRQYDKSDGEQLSARDLT 440
DB 308 PTGVITIGIETMNTSGAALITLNTSSGVTLTALNTTSGA-AQTVTAGAGQNLTA--TT 364
```

```

QY 441 WSYAALLTANRRNSVVPASWGETSASSVPGTCAATSAIGTYS-----SVTVSMPISVAT 496
DB 365 AAQANNVAVDGRANVTASVTSGTTTGG--ANSASGTVSIVSANSSTTTTGAIATV 422
QY 497 GGTTTATPTGSGSVTST-----SKTA-----TASKTSTTTRSG 531
DB 423 GGTAVTVAGTANAVNTTLTQADVTYTGNSSTTAVTVTQTAATAATAGATVAG 473

RESULT 26
US-09-142-648B-7
; Sequence 7, Application US/09142648B
; Patent No. 6210948
; GENERAL INFORMATION:
; APPLICANT: Smlt, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 6210948Belini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULOE
; FILE REFERENCE: 08106/002002
; CURRENT APPLICATION NUMBER: US/09/142,648B
; CURRENT FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: PCT/CA97/00167
; PRIOR FILING DATE: 1997-03-10
; PRIOR APPLICATION NUMBER: US 07/614,377
; PRIOR FILING DATE: 1996-03-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
; US-09-142-648B-7

Query Match      4.9%; Score 136; DB 3; Length 1026;
Best Local Similarity 23.2%; Pred. No. 0.009;
Matches 123; Conservative 73; Mismatches 225; Indels 110; Gaps 23;

QY 26 TLDSWLSNEATVATATILNNGADGAMVSGADSGIVASPTDNPDPFYTWTRDGLVVK 85
DB 28 TLDAAY-----ATQOTGGLSDAALTLTKLVNSTTAAVATQTYQ-----PFTGVADPSAAGLD 79
QY 86 TLVDLFRNGDTSLSLTIENTYISAQAIYQGISNPSGDLSSGAGLGEPRFNVDETAATYGSWG 145
DB 80 FLVDSTTN--TNDLN--DAYSKFAQENRFINFSINLATAGAGATAF--AAAYT----- 128
QY 146 RPQRDGPALATAMIGGQWLLDNGYTSATDVIWPLVRNDLSVYAOYMNQGYD-LMEE 204
DB 129 -----GVSYAQVATAYDKIIIGNAVATAAGVDVA-----AAVAFLSRQANIDYLTAF 175
QY 205 VNGSSEFTIAVQHRALVEGSAFATAVSGSCSCWCDGSAPELLCYLOSFWTGSFILANFDS 264
DB 176 VRANTPPTAADDIDLAVKALIGITILNAA-----TVSGIGGYATATAAMIN-DLS 224
QY 265 RSGKDANTLLG-SIHFPDPEAACDDSTFPQCSPPRALNHKEVDFSRSIYTLNDGLDSE 323
DB 225 DGLALSTNMAAGVNLFTAYPSSGVSGSTL-----SLTGTDTTLT--- 262
QY 324 AVAVAGRPEDTYNGNFWFLCTLAABQLYDALYQMDKQSLLEVTVSLDFKALYSDA- 382
DB 263 -----GTANNDTFVAGEVAGATTLVGGTLLSGG-----AGTIDV-LMWVQAAVATL 307
QY 383 ATGYSSSSSTYSIVDAVKTF--ADGFVSIIVETHAASNGMSRQYDKSDGEQLSARDLT 440
DB 308 PTGVITIGIETMNTSGAALITLNTSSGVTLTALNTTSGA-AQTVTAGAGQNLTA--TT 364
QY 441 WSYAALLTANRRNSVVPASWGETSASSVPGTCAATSAIGTYS-----SVTVSMPISVAT 496
DB 365 AAQANNVAVDGRANVTASVTSGTTTGG--ANSASGTVSIVSANSSTTTTGAIATV 422
QY 497 GGTTTATPTGSGSVTST-----SKTA-----TASKTSTTTRSG 531
DB 423 GGTAVTVAGTANAVNTTLTQADVTYTGNSSTTAVTVTQTAATAATAGATVAG 473
```

RESULT 27
US-08-728-470-10
Sequence 10, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Matzare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10
Query Match 4.8%; Score 133; DB 2; Length 1529;
Best Local Similarity 20.3%; Pred. No. 0.031;
Matches 119; Conservative 73; Mismatches 153; Indels 242; Gaps 28;
QY 15 TGLANVSK-RATL--DSWLSNEATVARTALNNIGADG-AWMSGADSGIV-----VASPS 66
DB 777 TGNINTTKRANVTLQADTISNSNGLKRLTLTGNISVEGLSLTGANANVGLSLAEDS 836
QY 67 T-----DNDPVEYTWTRDGLVLTVDLFRNGDTSLSTIBNIYISAQAIYOGISNPSG 120
DB 837 TTFGEASDNLNITGFT-----NNGTANI-----NIKGVKVLQG 871
QY 121 DLSSGGLGEPKRVNDETATYTSWGRPRQDGPALRAFAMIGFQMLDNGYSTATDIYW 180
DB 872 DINNKGL-----NI-----TTNASTGTOKTIINGNITTEKGLNLT 906
QY 181 PLVRNLSYVAOYMNQGVYL-----WEVNGSSFTTAVGHRALYEGSAFATAV 230
DB 907 KNIKADAEI-----QIGNISQKEGNTLTISSKVNITNQITL-----KAGVEG----- 949
QY 231 GSSCSWCSQCAPILCYLGSFTWGSFILANFDSRSQKDNLTLLGSIHFPDPAACDST 290

DB 950 GRS-----DSSER-ENANL---TIQTEKLKAGD--- 974
QY 291 PQPCSPRALANKEVVDSPRSITYTLN-DGLSDSEAVAVRYEDPYNPNPFLCTLAAR 349
DB 975 -----LNISGFNFAELTA----- 987
QY 350 EQLYDALYQMDKQSLVETVDSLDFEKKALYSDAATGYSSSSSYSDAVATFADGF- 408
DB 988 -----KNGS-----DLTI-----GNASGEMNAKKVTPKIVDS-KISTDGHN 1024
QY 409 VSIIVETHAASNGSMSEQYDKSDGDSIAPDLTWS-----YALLTANRRNSV 457
DB 1025 VILNISEVKTNSGSSNAGNDNSTGLITISAKDVTVNNVTSKTIINISAAAGNVTTKEGTTI 1084
QY 458 PASWGETSASSVPGTCAATSAIGTYSVTVTSMPSIV-----ATGCTTTATPTG-- 507
DB 1085 NATTGSVEVTAONGTITKNI-----TSQNVTVTAENLVTTENAVINATSGVINISTKTGDI 1141
QY 508 SCSVTST-----KTTATASKTSTTTRSGMSL 534
DB 1142 KGIBSTSGNNAVITASGNTLKXSNITGQDVTVTAADGALTTAGSTI 1188

RESULT 28
US-08-719-641-10
Sequence 10, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Matzare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10


```

RESULT 30
US-08-325-267A-2
/ Sequence 2, Application US/08325267A
/ Patent No. 5585271
/ GENERAL INFORMATION:
/ APPLICANT: MATARI, JUNJI
/ APPLICANT: TAKATA, YOSHIHIRO
/ APPLICANT: OGAMA, MASAHIRO
/ APPLICANT: PENTTILA, MERJA
/ APPLICANT: ONNELA, MAIJA-LEENA
/ APPLICANT: KERANEN, SIRKKA
/ TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
/ TITLE OF INVENTION: CONTAINING THEM
/ NUMBER OF SEQUENCES: 7
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
/ STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
/ CITY: ARLINGTON
/ STATE: VA
/ COUNTRY: USA
/ ZIP: 22202
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/325,267A
/ FILING DATE: 18-NOV-1994
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: JP PCT/JP94/00290
/ FILING DATE: 24-FEB-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: JP 38871/1993
/ FILING DATE: 26-FEB-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: OBLON, NORMAN F
/ REGISTRATION NUMBER: 24,618
/ REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 703-413-3000
/ TELEFAX: 703-413-2220
/ TELEX: 248855 OPAT UR
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1537 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-325-267A-2

Query Match 4.7%; Score 130.5; DB 1; Length 1537;
Best Local Similarity 20.0%; Pred. No. 0.051; 236; Indels 143; Gaps 27
Matches 119; Conservative 97; Mismatches 236; Indels 143; Gaps 27

QY 11 GLVCTGLANLVISKATLDSMTLSNEATVARTAIINNIGADGAWVSGADSGIVVASPSTNP 70
DB 769 GLIST-----TTEPW-TGRTTSTSTENTVTYTGNGOPTD--ETVIYVIRTPSSEC- 814
QY 71 DYFYTWTRD--SGLVAKTLVDLFRNGDTSLSTIENYI----SAQAIVQGISNP-SGDL 122
DB 815 --LVTTTTEPWTGTFTSTSTETTTTGNGVPTDETVIVIRTPSEGLISTTTEPW-TGRTTST 872
QY 123 SSGAGLGRKKNVDETAATGSGWRQQRD-----GPAIRATAMIGRCGLDLNGYTSSTAT 176
DB 873 TS-----TSTETTTITGNGOPTDETVIVIRTPSEGLISTTTEPW--TGRTTST 922
QY 177 D-----LWPLVRNDLSVYAQVNOTGYDLMEEV-----NGS-SFF 211
DB 923 EMTVTYTGNGVPTDETVIVIRTPSEGLISTTTEPWGTGFTTSTEVTTITGNGOPTDE 982
QY 212 TIAVQHRALVGSFAFAIAVAGSSCWDSQAPFIICLYLOSFWTGSFLIANFD---SSRSG 267

```

```

Db      983 TVIVIRPTSEGLISTTT-----EPMTGFTSTREMTVTGTG 1022
Qy      268 KDANTLLGSIHTPEPAACDSTFQPCSPRALNKHKEVDSFRSIYTLNDGLDSEAVAV 327
Db      1023 QPDEETVIVIRPTSEGLVTTT EPMTGFTSTSTEM-----STVGINGLPRDEFYIV 1076
Qy      328 GRVY----EDYYNGMPWFLCTLAAMEQLYDALYQMDKQGLSEVTVSLDFFALYSDA 382
Db      1077 VKPTPTTAISSLSSSSGQITTSSTSSRPITTPFPNGTVSVISSVSISSVTSSLFTSS 1136
Qy      383 ---ATGTVSSSSSTVSSIV-DAVTFPADGFVSIIVETHAASNGSMSEQYDKSDG----- 431
Db      1137 PVISSSVISSSTTTSTISIFSEBSKS-----SVTPTSSSTSGS-SESETSAGSVSSSSF 1189
Qy      432 -EQLSARDLTWVYAAALTANNRRNSVVPASWGETSASSVQTCATSAIGTVSSVTYTSW 490
Db      1190 ISSSSSKSPYVSSSLPLVTSATTS-----QETASSLPATTTKTSBQTLT--VTYVSC 1241
Qy      491 PSYATGCTT---TATPTGSGSVT-----STSKTATATASKSTTTR 529
Db      1242 ESHVCTESISPAIVSTATVTSVGTTEYTTWCPISTTETTKQTGTEQTTETTK 1296

RESULT 31
US-08-617-697-10
/ Sequence 10, Application US/08617697
/ Patent No. 5977336
/ GENERAL INFORMATION:
/ APPLICANT: Barenkamp, Stephen J
/ TITLE OF INVENTION: High Molecular Weight Surface Proteins
/ TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Shoemaker and Matzare, Ltd.
/ STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
/ STREET: Bldg. 1
/ CITY: Arlington
/ STATE: Virginia
/ COUNTRY: U.S.A.
/ ZIP: 22202-0286
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/617,697
/ FILING DATE: 01-APR-1996
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/302,832
/ FILING DATE: 05-OCT-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US PCT/US93/02166
/ FILING DATE: 16-MAR-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Berkstresser, Jerry W
/ REGISTRATION NUMBER: 22,651
/ REFERENCE/DOCKET NUMBER: 1038-557
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 415-0810
/ TELEFAX: (703) 415-0813
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1600 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-08-617-697-10

Query Match      4.7%; Score 130; DB 2; Length 1600;
Best Local Similarity 20.4%; Pred. No. 0.06;

```


APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Daneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5134:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1..673
SEQUENCE DESCRIPTION: SEQ ID NO: 5134:
US-09-107-532A-5134

Query Match 4.6%; Score 128.5; DB 4; Length 673;
Best Local Similarity 18.8%; Pred. No. 0.021;
Matches 113; Conservative 77; Mismatches 219; Indels 193; Gaps 26;

QY 18 ANVISKRATLDS-WLSNEATVARTAILNIGADGA-----WV----- 53
DB 85 ASVMTAQAQAVVESGM--GSSALSAQAPYNNFGIGKSYQGVYWDLEFLNNQVNNKEBP 142
QY 54 -----SGADSGIVVASPTDNPDYFTWTRDGLVLKTLVDLRNG-----DTSLLST 101
DB 143 RQYPSAESFNDAYVLRNTSGN-GYYAGTWKSTKSTYDTACTLGRVATDPQYAGK 201
QY 102 IENYISAQAIIVOGISNPSGDLSSAGLGEKFNVDATAVTSWGRFQDRG-----PAL 154
DB 202 LNNIIATYGLQDTPASGNAGGATG-----ANTGASGSGSSNGNNTSSSATST 256
QY 155 RATAMIGFQWLLDNGYSTAIDIV-WPLVRNDLSYVAQYWNQGYDLMEEVNGS----- 208
DB 257 TYVOSGDSVWGSINRFGITMDQLIOMNNIQQNFYPCQKLTIKG-----SQENGSSSTNS 312
QY 209 -----SFFITAVOHRALVGSAPATAVSSCSWCDSQAPBILCYLOSFTG----- 254
DB 313 GNNNTSSGAGTSGNGQTTGAKTYVOSGSV-WKISNDHGITMAQLIEMNNIKNNFVYPG 371
QY 255 -SFIILAFDSSRSRQKANTLIGSIHTFDEBACDSDTFQPCSPRALNKHVVDSFRSY 313
DB 372 QQLIVNGSSSNAGATGNTGNTGSSSSSSSSNT-----AGCTKY 414
QY 314 TLNDGLSDEAVAVGVPEDTYNGNPMFLCTIAAEBQLYDALYQWK-QGSLEVTDSL 372
DB 415 TVVAG-----DSVMSVANKGITWA-----QLIQNNNIQNNFIYPGQOL 453
QY 373 DFFKALYDAATGTYSSSSSTYSIYDAVKTFPADGVSIVETHAANGSGSEQ-YDKSD 430
DB 454 TVNSNGSTVSTSGNTSGSTNTNS-----SSVNTGSSSSGTYTVKA 495
QY 431 GEOLSRDLTWSYA-----ALLTANRRNSVVPASWGETSASSVPGTCAATSAIGTY 482
DB 496 GBSV-----WSVANNGGITMDLIEMNNIKNNFI-----YPG----- 527
QY 483 SSVTVTSPBIVATGTTTATPYGSGSVTSTK-----TTATASKTST-----TTRSGM 532
DB 528 -----QQLIVKGGTSAAN-----TNTGSTTSAKPMTPNTSATSTSTGNTMYTVKAGS 576

QY 533 SL 534
DB 577 SV 578

RESULT 34
US-08-700-651-5
Sequence 5, Application US/08700651B
Patent No. 6015882
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
CURRENT FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751
EARLIER FILING DATE: 1995-04-03
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 5
LENGTH: 1721
TYPE: PRT
ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Query Match 4.6%; Score 128; DB 3; Length 1721;
Best Local Similarity 22.8%; Pred. No. 0.1;
Matches 50; Conservative 31; Mismatches 102; Indels 36; Gaps 5;

QY 331 PEDTYNN-----GNPWFELCTIAAEBQLYDALYQWKXGSLVTDVSLDFFKALYSDA 383
DB 82 PTDPSYNCPEFNPVTGVLVSRSTGKTIPTNYAGV--RNETKTEPSSANTYAGVRSNE 138
QY 384 TGTYSSTSSSTYSIYDAVTFADGVSIVETHAANGSGMS-RO-----YDKS 429
DB 139 KTEPSSANTNPLVDP-----KINAPCNSNSFQGGQIFDMGSKVYIPYTK 186
QY 430 DEOLSRDLTWSYAALLTANRRNSVVPASWGETSASSVPGTCAATSAIGTYSVTST 489
DB 187 VGKKTITTT 246
QY 490 WPSIVATGTTTATPYGSGSVTSTKTTATASKTSTTT 528
DB 247 TT 285

RESULT 35
US-08-928-361B-6
Sequence 6, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIRKA
STREET: 385 Sherman Avenue, Suite 6
STATE: CA
CITY: Palo Alto
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: VERNY, HANA
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1 (HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
FAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1721 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-6

Query Match 4.6%; Score 128; DB 3; Length 1721;
Best Local Similarity 22.8%; Pred. No. 0.1;
Matches 50; Conservative 31; Mismatches 102; Indels 36; Gaps 5;
QY 331 PETHYNN-----GNPFLCTLAABQLYQMDKQSLVDVSLDFKALYSDDA 383
DB 82 PTDPYSCNCPNPTGNLVSRTGKTIPTNYAGVY---RSNETTTPSANTYAGVYRSNE 138
QY 384 TGYSSSSSTYSIVDAVKTADGFSIVETHAASGMS-EQ-----YDKS 429
DB 139 TKTEPSANTNPLVDP-----KINAPCNSNPFEGQIFDMGSKYIPIYTKC 186
QY 430 DGEQLSARDLWTSYAALLTANNRRNSVPAWGETSASSPGTCATSAIGTYSVTVS 489
DB 187 VGVKHTT 246
QY 490 WPSIVATGTTTATPTGSGSVTSKTTATASKTSTTT 528
DB 247 TT 285

RESULT 36
US-09-588-995A-6
Sequence 6, Application US/09588995A
Patent No. 6514697
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: BARNES, DEBRA A.
APPLICANT: NELSON, RICHARD C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND
TITLE OF INVENTION: ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM
FILE REFERENCE: 480.19-5
CURRENT APPLICATION NUMBER: US/09/588,995A
CURRENT FILING DATE: 2000-06-06
PRIOR APPLICATION NUMBER: 08/827,171
PRIOR FILING DATE: 1997-03-27
PRIOR APPLICATION NUMBER: 08/928,361
PRIOR FILING DATE: 1997-09-12
PRIOR APPLICATION NUMBER: 08/700,651
PRIOR FILING DATE: 1996-08-14
PRIOR APPLICATION NUMBER: 08/415,751
PRIOR FILING DATE: 1995-04-03
NUMBER OF SEQ ID NOS: 115
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 1721
TYPE: PRT

ORGANISM: Cryptosporidium parvum
US-09-588-995A-6
Query Match 4.6%; Score 128; DB 4; Length 1721;
Best Local Similarity 22.8%; Pred. No. 0.1;
Matches 50; Conservative 31; Mismatches 102; Indels 36; Gaps 5;
QY 331 PETHYNN-----GNPFLCTLAABQLYQMDKQSLVDVSLDFKALYSDDA 383
DB 82 PTDPYSCNCPNPTGNLVSRTGKTIPTNYAGVY---RSNETTTPSANTYAGVYRSNE 138
QY 384 TGYSSSSSTYSIVDAVKTADGFSIVETHAASGMS-EQ-----YDKS 429
DB 139 TKTEPSANTNPLVDP-----KINAPCNSNPFEGQIFDMGSKYIPIYTKC 186
QY 430 DGEQLSARDLWTSYAALLTANNRRNSVPAWGETSASSPGTCATSAIGTYSVTVS 489
DB 187 VGVKHTT 246
QY 490 WPSIVATGTTTATPTGSGSVTSKTTATASKTSTTT 528
DB 247 TT 285

RESULT 37
5457037-5
Patent No. 5457037
APPLICANT: TOGNONI, ANGELO;CARRERA, PAOLO;CAMERINI, BARBARA;
GALLI, GIULIANO;LUCCHESI, GIUSEPPE;GRANDI, GUIDO;DI GENNARO, CARLO
TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
NUMBER OF SEQUENCES: 7
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/1,797
FILING DATE: 08-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 749,621
FILING DATE: 19-AUG-1991
APPLICATION NUMBER: 224,114
FILING DATE: 25-JUL-1988
SEQ ID NO:5
LENGTH: 751
5457037-5
Query Match 4.6%; Score 127.5; DB 6; Length 751;
Best Local Similarity 21.8%; Pred. No. 0.03;
Matches 146; Conservative 67; Mismatches 213; Indels 245; Gaps 38;
QY 45 NIGADGAWMSGADSGI-----VWASPTDNPDIPTWTDSGLVLTVDLFRNGDTSL 99
DB 139 NVPASGASYRTTDSGIYAPKGVVLPSTOSTGKTPRAQKDDVIVHVRGFTEDTSLP 198
QY 100 STIE-----NTISAQAI-----VQGISNPSGL--SSGAG-----LGPKEF 134
DB 199 AQYRGTYGAGLAKASILASIGTAYVEFLVQETQNDANDVPSNDANQYWGNTENYS 258
QY 135 VDET-AYTSGWRPORDGPALRATAMI-----GF-----GQWLLDNGYTS 173
DB 259 PDRRYAYNNKAAG-----GPTAEQAMVQAFHNAIGIKVMDVYVNHAEAGTWTSSD--PT 311
QY 174 TATDIWPLVRDLSTVAQYMNQTYDLMEEVNGSSFFIT-AVQHAIYEGSAF----- 226
DB 312 TATVEMRGIDNTTYELTSGNQFYD--NTGIGANFNNTYVAQNIYDLSLAYMANTWG 369
QY 227 -----ATAVSSC--SWCDSQAPE-----ILCYLOSF----- 251
DB 370 VDGFRDLASVLSNCLNGAYTASAPNCPCNGCYNFPDAADSNVAIINILREFTYRPAAGS 429
QY 252 -----WT--GSFILLNF-----DSSRSQ-----KXANTILGS 276
DB 430 GDLFAPBPAIIGNSYQLGSPQGWSEWNGLFRDLSLRQONELGSMITIVYQANDPFGS 489
QY 277 IHTFDEAACDSTFPQCPSPALANHKEYVD--SFNSITYLNDGLDSEAVAVGRIPEDT 334

Db 490 SNLFQSSGR-----SPMNSINFIDVHDGWTLDKDVYSCN--GANNQAMPYG--PSDG 537
Qy 335 YVNGNFWPLCTLLAAEQLYDALYQWPKQSGLEVTDVSLDFPKALYSDA-----TGT--- 386
Db 538 GISTN-----YSWD-QQMSAGTGAVDQRRARAGMAMFEMLSAGTPLM 579
Qy 387 -----YSSSSS-----TYSISIVAVK--TFADGFVSIIVETHAANGSMSE 424
Db 580 QGGDEVILRLQCNNAYNLDDSANMLTYSWTTDQSNFYFAQRLIAFRKAHPLRPS--- 636
Qy 425 QYDKSDGEQLSARDLTWMSYAAALLTANNRRNSVVPASGETS-----ASSVGTCAATSAG 480
Db 637 -----SWYSGSQLTWQPSGAVADSN-----YMNNTSNVIAIYAING-----PSLG 677
Qy 481 TYSSVTV--TSPSIVATGCTTTATPTGSG--SVTST-----SKTTATASKT----- 524
Db 678 DSNISIVAYNGWSSSVT---FTLPAPPSGTQWYRVTDTCMDNDGASTFVAPGSETLIGA 734
Qy 525 -STTTRSGMSL 534
Db 735 GTTYGCGQGS 745

RESULT 38

US-09-346-237-7
; Sequence 7, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Biagard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629-200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 776
; TYPE: PRP
; ORGANISM: Pseudomonas species SMP1
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)
; OTHER INFORMATION: Isoamylase
US-09-346-237-7

Query Match 4.6%; Score 127.5; DB 3; Length 776;

Best Local Similarity 21.8%; Pred. No. 0.031;
Matches 146; Conservative 67; Mismatches 213; Indels 245; Gaps 38;

Qy 45 NIGADADAWYSGADSGI-----VVASPTQNDPQFYMTWTRSGVLTIVLDFNGSTSL 99
Db 164 NVFASASASTRTDSTGIYAPKGVLPVSTOSTGTFKPRAQKDVYEHVHVGFTEDJSLP 223
Qy 100 STIE-----NYISAQAI-----VQGISNPSGDL--SSGAG-----LGEPEK 134
Db 224 AQYRGTYGAGLKLASGLASGLVAVFLPQGEIQONANDVPKSDANQNWGIMTENYS 283
Qy 135 VDET-AYTSWGPQSDGALRATAMI-----GF-----GQMLDNGYTS 173
Db 284 PDRRYVYNKAAG-----GPTAEPOAWQAFHNAGIKVYMDVYVNHRTAEGGTWTSSD--PT 336
Qy 174 TATDIWPLVRNDLSVVAQWNTQGYDLMEVNGSSFFTL-AVQHRALVEGSAF----- 226
Db 337 TATITWGRIGDNTTYIELTSGNQFYD--NTGIGANFNNTVYAQNLIVDSGLAYMANTWG 394
Qy 227 -----ATVAVSSC--SWCDSQAP-----ILCYLOSF----- 251

Db 395 VDFRFDLASVLGNSCLNGAYTASAPNCBNGYFNDAADSNVAINRLIEFTVPPAGGS 454
Qy 252 -----WT-----GSFILANF-----DSSRSG-----KQANTLLGS 276
Db 455 GUDLFAEPALIGNSYQLQGFPGWSENNGLPRDLSRQONELGSMITTYTOANDPSSGS 514
Qy 277 HTFDEAACDSTFQPCSPRALANKEVVD--SFRSIYTLNDGLDSEAVAVGRYPEDT 334
Db 515 SNLFQSSGR-----SPMNSINFIDVHDGWTLDKDVYSCN--GANNQAMPYG--PSDG 562
Qy 335 YVNGNFWPLCTLLAAEQLYDALYQWPKQSGLEVTDVSLDFPKALYSDA-----TGT--- 386
Db 563 GISTN-----YSWD-QQMSAGTGAVDQRRARAGMAMFEMLSAGTPLM 604
Qy 387 -----YSSSSS-----TYSISIVAVK--TFADGFVSIIVETHAANGSMSE 424
Db 605 QGGDEVILRLQCNNAYNLDDSANMLTYSWTTDQSNFYFAQRLIAFRKAHPLRPS--- 661
Qy 425 QYDKSDGEQLSARDLTWMSYAAALLTANNRRNSVVPASGETS-----ASSVGTCAATSAG 480
Db 662 -----SWYSGSQLTWQPSGAVADSN-----YMNNTSNVIAIYAING-----PSLG 702
Qy 481 TYSSVTV--TSPSIVATGCTTTATPTGSG--SVTST-----SKTTATASKT----- 524
Db 703 DSNISIVAYNGWSSSVT---FTLPAPPSGTQWYRVTDTCMDNDGASTFVAPGSETLIGA 759
Qy 525 -STTTRSGMSL 534
Db 760 GTTYGCGQGS 770

RESULT 39

US-08-797-366-3
; Sequence 3, Application US/08797366
; Patent No. 5853702
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Christgau, Stephan
; APPLICANT: Halkier, Torben
; APPLICANT: Shuster, Jeff
; APPLICANT: Fuglsang, Claus Crone
; TITLE OF INVENTION: Penicillium Purpurogenum Mutanases
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSER: No. 58537020 No. 5853702disk of No. 5853702ch America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; APPLICATION NUMBER: US/08/797,366
; FILING DATE: 09-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4593-200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-797-366-3

Query Match 4.6%; Score 127; DB 2; Length 630;
Best Local Similarity 21.5%; Pred. No. 0.025;
Matches 131; Conservative 77; Mismatches 229; Indels 172; Gaps 32;

```
QY 2 SFRSLALSGVCTGLA--NVISKRAITLDSMLSEAVIA---RTALINNIGAD--GAWV 53
DB 6 AFATATSAIIAACSLPSDSMVSRRSTSDRLVFHFVWGVLSRPTS--ASDYADMQAKA 64
QY 54 SGADS-GIVVASPSTNDPFIYTWTRDSGLVLTVDLFRNGDTSLSTIENTISAQAIY 112
DB 65 YGIDAFALNIGTDFPSQOQAGYAESANNDMKVFISF---DFNMWSTGATEIGKIA 120
QY 113 QGISNPSGDLSSGAGLCEPKFENVDETAYTGSWGRPQRDPALRATAMIGFGWMLDNGYT 172
DB 121 QYGSLLPG-----OLMYDDKIFVSSFAGDGVVALKSA--GGNVFAPAFPH 166
QY 173 STATDIY-----WPLVRNDLSVAQVMNQTGYDLMEVVGSSFFITAVQHRALVE 222
DB 167 SYGTDLSDVDGLLNMGMFSGNKKAPTA-----GANVTVREGDEEYIT-ALDGKPYIA 219
QY 223 GSA--FATAVSSCSCWCDQA--PEILCYLOSFWTGSFIANF-----DSSR 265
DB 220 PASWFEFTHGPEVTYSKNWVFPDDLFXQR--WMD---LNLGPOFLEVVTWMDYGSQY 275
QY 266 SGKDANTLLGSIHTFDEPA--ACD--DSTFQPCSPALANHKRVDSFRSIYTLNDGL- 319
DB 276 VGP-----LNSPHTDDSSRWANMPHDGWLDAKPYIAAFH--DGATSLSSSYTTEDQLI 329
QY 320 -----SDSEAVAVGRRPEDT--YYNGNP--WFLCTLLAAEQLYDALYOWDK 361
DB 330 YWRPQRLMDCATDTCWVAAN--NDTGNVFEGRNGW-----ESMEDAV----- 373
QY 362 QGSLEVTDSLDPEFKALYSDAATGTYSSSSTYSIVDAVTFADGVSIETHAASNGS 421
DB 374 -----FVALLQSACTYQVTSQPNY-----ETDAPAGA 402
QY 422 MSEQYD-----KSDEQL-----SARDLT-----WSYALLTANNRRNSV 457
DB 403 SAFQVPMGFQPSFSLSRDGETVLSGLKDIIIDCLGIVFNAYVGLPA--TFSDPLE 461
QY 458 PASWGEFSSASVPTCAATSAIGTYSVTVSWPSIVATGTTTATPTGSGVTSYST 517
DB 462 PPSLNAPSEGLKSTCATSPSLG---LTSTTPPETIPTGITGSAITGAATTSTIST 517
QY 518 TATASKTST 526
DB 518 TSTISTST 526
```

RESULT 40
US-08-956-268-3
Sequence 3, Application US/08956268
Patent No. 5874275
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Christgau, Stephan
APPLICANT: Halkier, Torben
APPLICANT: Shuster, Jeff
APPLICANT: Fuglsang, Claus Crone
TITLE OF INVENTION: Penicillium Purpureogenum Mutanaes
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58742750 No. 5874275disk of No. 5874275th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,268
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/797,366
FILING DATE: 09-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4593.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-956-268-3
```

Query Match 4.6%; Score 127; DB 2; Length 630;
Best Local Similarity 21.5%; Pred. No. 0.025;
Matches 131; Conservative 77; Mismatches 229; Indels 172; Gaps 32;

```
QY 2 SFRSLALSGVCTGLA--NVISKRAITLDSMLSEAVIA---RTALINNIGAD--GAWV 53
DB 6 AFATATSAIIAACSLPSDSMVSRRSTSDRLVFHFVWGVLSRPTS--ASDYADMQAKA 64
QY 54 SGADS-GIVVASPSTNDPFIYTWTRDSGLVLTVDLFRNGDTSLSTIENTISAQAIY 112
DB 65 YGIDAFALNIGTDFPSQOQAGYAESANNDMKVFISF---DFNMWSTGATEIGKIA 120
QY 113 QGISNPSGDLSSGAGLCEPKFENVDETAYTGSWGRPQRDPALRATAMIGFGWMLDNGYT 172
DB 121 QYGSLLPG-----OLMYDDKIFVSSFAGDGVVALKSA--GGNVFAPAFPH 166
QY 173 STATDIY-----WPLVRNDLSVAQVMNQTGYDLMEVVGSSFFITAVQHRALVE 222
DB 167 SYGTDLSDVDGLLNMGMFSGNKKAPTA-----GANVTVREGDEEYIT-ALDGKPYIA 219
QY 223 GSA--FATAVSSCSCWCDQA--PEILCYLOSFWTGSFIANF-----DSSR 265
DB 220 PASWFEFTHGPEVTYSKNWVFPDDLFXQR--WMD---LNLGPOFLEVVTWMDYGSQY 275
QY 266 SGKDANTLLGSIHTFDEPA--ACD--DSTFQPCSPALANHKRVDSFRSIYTLNDGL- 319
DB 276 VGP-----LNSPHTDDSSRWANMPHDGWLDAKPYIAAFH--DGATSLSSSYTTEDQLI 329
QY 320 -----SDSEAVAVGRRPEDT--YYNGNP--WFLCTLLAAEQLYDALYOWDK 361
DB 330 YWRPQRLMDCATDTCWVAAN--NDTGNVFEGRNGW-----ESMEDAV----- 373
QY 362 QGSLEVTDSLDPEFKALYSDAATGTYSSSSTYSIVDAVTFADGVSIETHAASNGS 421
DB 374 -----FVALLQSACTYQVTSQPNY-----ETDAPAGA 402
QY 422 MSEQYD-----KSDEQL-----SARDLT-----WSYALLTANNRRNSV 457
DB 403 SAFQVPMGFQPSFSLSRDGETVLSGLKDIIIDCLGIVFNAYVGLPA--TFSDPLE 461
QY 458 PASWGEFSSASVPTCAATSAIGTYSVTVSWPSIVATGTTTATPTGSGVTSYST 517
DB 462 PPSLNAPSEGLKSTCATSPSLG---LTSTTPPETIPTGITGSAITGAATTSTIST 517
QY 518 TATASKTST 526
DB 518 TSTISTST 526
```

RESULT 41

US-09-717-364A-15
; Sequence 15, Application US/09717364A
; Patent No. 6663872
; GENERAL INFORMATION:
; APPLICANT: Pitkovski, Jacob
; APPLICANT: Muelem, Margalit
; APPLICANT: Koren, Ziv Re1
; APPLICANT: Krippe1, Simcha
; APPLICANT: Shmueli, Esther
; APPLICANT: Perez, Yifat
; APPLICANT: Guter, Bezalel
; APPLICANT: Gaili1, Galad
; APPLICANT: Michael, Amnon
; APPLICANT: Goldberg, Doron
; TITLE OF INVENTION: HEMORRHAGIC ENTERITIS VIRUS DNA SEQUENCES, PROTEINS ENCODED THERE
; TITLE OF INVENTION: VARIOUS USES THEREOF
; FILE REFERENCE: 1567/63655
; CURRENT APPLICATION NUMBER: US/09/717,364A
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: IL124567
; PRIOR FILING DATE: 1998-05-20
; PRIOR APPLICATION NUMBER: PCT/IL9900268
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patent version 3.1
; SEQ ID NO 15
; LENGTH: 906
; TYPE: prf
; ORGANISM: hemorhagic enteritis virus
US-09-717-364A-15

Query Match 4.6%; Score 127; DB 4; Length 906;
Best Local Similarity 21.8%; Pred. No. 0.044;
Matches 97; Conservative 58; Mismatches 155; Indels 134; Gaps 21;

QY 155 RATAMIGFQMLNDNGYSTAIDVMPVLR-----NDLSYVAQYMNQGYDLM 202
DB 83 RTIAVGDGR-VLDMG--STYFDIRGNIDRGPSFKYGTAVYPLAPRGAQFNMI----- 134
QY 203 EEVNGSSFFTIAGVQHALVEGSAFATAVSSCGSCWCSQAPEILCYQSTWTSFIILAND 262
DB 135 KTVGGKTYLT-----AQATKFFSTSGNGCAANTEASSFTNLVPSPTGS----- 179
QY 263 SRSRGKANTLLGSHITPPEAACDSTFQCPSPRALNHNKEVVDSEFRSITYLINDLSDS 322
DB 180 -----AESSEDP-----TEGASCRAL-----TL--GSSVT 203
QY 323 EAVAVGRYPEDTYNGN---PWFLCTIAAEQLYDALYOMDKOGLLEV- DVSIDFFKAL 378
DB 204 DATCGAVYPIQNGANSILRP-----SVTPDKRPADA-----GKSGSVTCTALICCDNVTVQ 255
QY 379 YSDAATGTYSSSSSTYSIVDAVK--TFADGFSIV-ETHAASNGM-SEQYDKSDGEOL 434
DB 256 YPDRIVAVDSYDKLATRMGNRINYIGFRDNFGLMYDNGAHSGLATETGDIINVEOL 315
QY 435 SARDLTWSYALTLTANRRNSVVPASWGET-----SASSVPGCATSAI 479
DB 316 QBRNTEISIQWMLADLSNRNHY-SQMNQAVDDYDLNVRVLTNIGYEEGPGCYPTGCM 374
QY 480 GTYSSVTVTSWPSIYATGTTTAPTP-----GSGSVTST----- 514
DB 375 GNYPN-TWMSVGLVLDNNGTATATTNTVAVVGFGSVPTMEINVOAYLQCKMYANIAIY 433
QY 515 -----SKTATATASKSTTTRSGMS 533
DB 434 LPDKYKKAIGTSETDPTTYSYMN 457

RESULT 42
US-09-346-237-4
; Sequence 4, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:

; APPLICANT: Biesgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 776
; TYPE: prf
; ORGANISM: Pseudomonas amylocleromosa
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)
; OTHER INFORMATION: Isoamy1ase
US-09-346-237-4

Query Match 4.6%; Score 126.5; DB 3; Length 776;
Best Local Similarity 21.8%; Pred. No. 0.038;
Matches 146; Conservative 67; Mismatches 213; Indels 245; Gaps 38;

QY 45 NIGADGAWISGADSGI-----VVASPSTNDPDIYFTWTRDSGLVLTLYDLFRNGDTSL 99
DB 164 NVFASGASRYRTDTSIGYAPKGVLVSTQSTKPRACKDDVIEVHVGFTEDQTSIP 223
QY 100 STIE-----NYISAQAI-----VQGISNPSGL--SSGAG-----LGSPKFN 134
DB 224 AQYRGTYAGAKASTIASLGYTAVEFLVQETQDADADVENSQANQYKMGTYNTYS 283
QY 135 VDET-AYTGSKWRPORDGPALRATMI-----GF-----GQMLNDNGYS 173
DB 284 PRRRYAVNKAAG-----GPTAEQAVVQAFHNAQIKVYMDVYVYNNHTAEGTWTSSD--PT 336
QY 174 TATDIWPLVRNDLSYVAQYMNQGYDLMEEVNGSSFFTI-AVQHALVEGSAF----- 226
DB 337 TATYISWRGLDVAITYELTSGNQFYD--NTGIGANFNFTYNTVAQNLIVDSLAWMANTWG 394
QY 227 -----ATAVGSSC--SMCSQAPB-----ILCYLOSF----- 251
DB 395 VDGFRPDLASVYGNCLNCAITYASAPNCNGYNDPADSNVAIRILREFYVTPAAGS 454
QY 252 -----WT--GSEFIILNF-----DSRSG-----KDANTLLGS 276
DB 455 GDLFAEPWALIGNSYQGGFQGWSEWNGLFPRDSLQKQNELGSMWITIVIQDANDFGS 514
QY 277 HTFPPEAACDSTFQCPSPRALNHNKEVVD--SRSITYLNDGLSDSEAVAVGRYPEDT 334
DB 515 SNLFQSSGR-----SPWNSINFIDVHDQMTLKDVSCH--GANNSSQAMPYG--PSDG 562
QY 335 YNNGNPMFLCTIAAEQLYDALYOMDKOGLLEVTVDSIDFFKALYSDA-----TGT 386
DB 563 GTSTN-----YSMD-QGMSAGTAANDQRPARTGMAFEMLSACTPYLM 604
QY 387 -----YSSSSS-----YSSSIVDAVK--TFADGFSIVETHAASNGMSB 424
DB 605 QGDEYLTFLQCNNAVYNDSSANWLTYSWTTDQSNFYFAQRLAFRAHALARS----- 661
QY 425 QYDKSDGEQLSARDUTWSYALTLTANRRNSVVPASWGETS-----ASSVPGCATSAIG 480
DB 662 -----SWYSGSGLTWYQPSGAVADSN-----YMNNTSYAIAIAYING-----PSLG 702
QY 481 TYSSVTV--TSPSIYATGTTTAPTPSG--SVTST-----SKTATATASKT----- 524
DB 703 DSNSTIYAVNGSSSVT---FTLPAPPSGTQMYRYTDTCDMDAGASTFVAPBSFTLIGA 759
QY 525 -STTRSGMSL 534
DB 760 GTTYGCGGSL 770

Qy 184 RNDLSVYAGYWNQGYDLWEEVNGSFFITAVQHRAIVBGSAPATAVSSC-SWCDQAP 242
 Db 141 -----TGFFLPQ-TGSYTFKAT-----VDSALISVGATAFNCQAQQP 181
 Qy 243 EILCYLQSF-----WTGS-----FILANFDSRSRGKANTLLGSIHTFDPBAAC 286
 Db 182 PITS--TNFTIDGIKPMGSLPPNIEGTVMYAGYYPMKVYSNAV--SWGTLPISVTL 237
 Qy 287 DSTFQPCSPBALANHKEVVDSPRS-IYTLNDGLSDSEAVAVGRYPEDITY-----NG 338
 Db 238 PGCT-----TVSDPEGVYISFDDDLQSNCV---PDPSNVAVSTTTT 280
 Qy 339 NPWFLLCTLAABQLYDALYQMDKQSLLEVTVSLDFFKALYSDAATGYSSSSSTYSIY 398
 Db 281 BPW-----TGFTSTSTEMTVT 298
 Qy 399 DAVKTFADGFVSIVETHAASNGSMSEQYDKSGEQLSARDLWVSVALLTANNR--NS 455
 Db 299 GNGVPTDETVIVIRT-PTSEGLISTTBEPWTGFTS---TSTEVTITGTNGQPTDET 353
 Qy 456 VV-----PASMGETSASSVPGTCAATSAIGTVSSVTWSPSIIVATGTTT-----TATP 505
 Db 354 VIVIRPTSEGLISTTBEPWT-----GFTS-TSTEMTVTGTNGQPTDETVIVIRTP 405
 Qy 506 TSGSVYTSKT-----TATASKTSTTT 528
 Db 406 TSEGLVTTTTEPWTGTSTSTEMSTVT 433

Search completed: June 28, 2004, 07:45:58
 Job time : 28 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:39:37 ; Search time 60 Seconds
(without alignments)
2514.673 Million cell updates/sec

Title: US-10-038-723-2

Perfect score: 2771

Sequence: 1 MSFRLSLALSGIVCTGLANV.....SKTTATASKSTYTRSGMSL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database: A_Geneseq_290a04:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2771	100.0	534	2	AAV18090 Truncated
2	2771	100.0	534	2	AAV23338 Aspergill
3	2771	100.0	534	3	AAAB03450 Aspergill
4	2771	100.0	534	3	AAV77740 A. niger
5	2771	100.0	534	4	AAAB48171 A. niger
6	2771	100.0	534	4	AAAB61904 A. niger
7	2742	99.0	640	4	AAV77741 A. niger
8	2742	99.0	640	4	AAAB61905 A. niger
9	2683.5	96.8	639	1	AAAP0212 Sequence
10	2658	95.9	630	1	AAAB1876 Sequence
11	2658	94.8	616	3	AAAB15176 Aspergill
12	2655	94.7	616	2	AAAB55979 Aspergill
13	2655	94.7	616	3	AAAB5180 Aspergill
14	2615	94.4	616	2	AAAB55976 Aspergill
15	2615	94.4	616	3	AAAB5178 Aspergill
16	2614	94.3	616	3	AAAB15184 A. awamori
17	2612	94.3	616	2	AAAB55980 Aspergill
18	2604	94.0	616	2	AAAB5181 Aspergill
19	2604	94.0	616	2	AAAB55977 Aspergill
20	2600.5	93.8	621	2	AAAB5179 A. awamori
21	2600.5	93.8	621	2	AAAB55978 Aspergill
22	2589.5	93.5	621	3	AAAB15183 A. awamori
23	2584.5	93.3	621	3	AAAB15182 A. awamori
24	2523	91.1	1095	6	ABP96630 Alpha-amy
25	2468	89.1	471	4	ABM00045 AMG SEQ I

26	1850.5	66.8	631	6	ABBB0181 A. fumiga
27	1770	63.9	624	4	AAAB48170 Thermococ
28	1689	61.0	618	2	AAV23339 Talacomy
29	1674.5	60.4	591	2	AAV23337 Talacomy
30	1615.5	58.3	630	5	AAAB51596 Thielavia
31	1577	56.9	626	2	AAV71034 N. crassa
32	1425.5	51.4	581	3	AAAB18823 Amino act
33	1425.5	51.4	581	5	AAV79444 Fusarium
34	1425.5	51.4	581	7	ABO01925 Fusarium
35	1408.5	50.8	704	6	ABBB0170 A. fumiga
36	1380.5	49.8	620	6	ABBB0174 A. fumiga
37	1321	47.7	616	6	AAAB30155 Glucoamyl
38	725	26.2	579	6	ABBP6632 Rhizopus
39	724	26.1	604	1	AAAP60359 Glucoamyl
40	641.5	23.2	624	2	AAV7674 Glucoamyl
41	615.5	22.2	497	1	AAV70572 Glucoamyl
42	600.5	21.7	918	1	AAAP60723 Sequence
43	374	13.5	293	1	AAAP70183 Sequence
44	181	6.5	1588	7	ADC01413 Enterohae
45	172	6.2	32	2	AAAB22343 His(184)
46	170	6.1	32	2	AAAB22338 His(177)
47	169	6.1	32	2	AAAB22341 Ala(182)
48	169	6.1	32	2	AAAB22342 Lys(183)
49	169	6.1	32	2	AAAB22340 His(183)
50	164.5	5.9	2344	4	AAU37120 Staphyloc
51	164	5.9	32	2	AAAB22345 Lys(183)
52	164	5.9	32	2	AAAB22346 Ala(182)
53	164	5.9	2016	6	AAAB36891 piecreur
54	163	5.9	32	2	AAAB22339 Arg(178)
55	161	5.8	32	2	AAAB22344 Ala(182)
56	161	5.8	1063	5	ABBB54168 Lactococ
57	160	5.8	35	2	AAV23334 Peptide d
58	159.5	5.8	2261	6	ABJ18914 Pathogen
59	159.5	5.8	2271	6	ABW72734 Staphyloc
60	159.5	5.8	2283	6	ABP56876 Staphyloc
61	159	5.7	2271	6	ABU16000 Protein e
62	158	5.7	2870	3	AAV95559 Caenorhab
63	158	5.7	3178	3	AAV95556 Caenorhab
64	156.5	5.6	2137	5	ABP39618 Staphyloc
65	155.5	5.6	596	6	ABBS9339 Human sec
66	154.5	5.6	1031	4	AAU35339 Enterococ
67	152	5.5	45	6	ABBP6634 Raw-starc
68	151.5	5.5	1099	6	ABU020296 Protein e
69	147.5	5.3	4440	6	ABU88256 Novel hum
70	147.5	5.3	4440	6	ABU90135 Novel hum
71	147.5	5.3	4440	6	ABU96437 Novel hum
72	147.5	5.3	4440	6	ABU99046 Novel hum
73	147.5	5.3	4440	6	ABU98261 Novel hum
74	147.5	5.3	4440	6	ABU91967 Novel hum
75	147.5	5.3	4440	6	ABU85271 Novel hum
76	147.5	5.3	4440	6	ABU80410 Novel hum
77	147.5	5.3	4440	6	ABU88961 Novel hum
78	147.5	5.3	4440	6	ABU06457 Novel hum
79	147.5	5.3	4440	6	ABU95517 Novel hum
80	147.5	5.3	4440	6	ABU95207 Novel hum
81	147.5	5.3	4440	6	ABU90755 Novel hum
82	147.5	5.3	4440	6	ABU93917 Novel hum
83	147.5	5.3	4440	6	ABU86191 Novel hum
84	147.5	5.3	4440	6	ABU82046 Novel hum
85	147.5	5.3	4440	6	ABU07907 Novel hum
86	147.5	5.3	4440	6	ABU94227 Novel hum
87	147.5	5.3	4440	6	ABO00410 Novel hum
88	147.5	5.3	4440	6	ABU87111 Novel hum
89	147.5	5.3	4440	6	ABU91352 Novel hum
90	147.5	5.3	4440	6	ABU90445 Novel hum
91	147.5	5.3	4440	6	ABU97036 Novel hum
92	147.5	5.3	4440	6	ABO05232 Novel hum
93	146.5	5.3	4469	6	ABU42529 Protein e
94	146.5	5.3	1439	5	ABU16643 Protein e
95	144.5	5.2	614	6	ABBS4071 Lactococ
96	144.5	5.2	738	7	ADC01840 C. albica
97	143.5	5.2	596	3	AAV99408 Human PRO
98	143.5	5.2	596	4	AAAB6157 Protein o

99 143.5 5.2 596 4 AAU29178
100 143.5 5.2 596 4 AAB87575

Aau29178 Human PRO
Aab87575 Human PRO

ALIGNMENTS

RESULT 1

AA18090
ID AA18090 standard; protein; 534 AA.

XX AA18090;

XX 06-AUG-1999 (first entry)

XX Truncated A. niger glucoamylase G1 protein sequence.

XX Glucoamylase; saccharification; liquefied starch solution; cyclodextrin;
XX enzymatic saccharification; monosaccharide production; oligosaccharide;
XX dextrose; trehalose; isomaltoligosaccharide; maltoligosaccharide.

XX Aspergillus niger.

XX MO9927124-A1.

XX 03-JUN-1999.

XX 23-NOV-1998; 98WO-US024871.

XX 26-NOV-1997; 97US-00979673.

XX 30-JUN-1998; 98US-00107657.

XX (NOVO) NOVO-NORDISK AS.

XX (STAL) STALEY MFG CO A E.

XX liaw GC, Pedersen S, Hendriksen HV, Svendsen A, Nielsen BR,

XX Nielsen RI;

XX WPI; 1999-357849/30.

XX N-PSDB; AAX76983.

XX Saccharification of liquefied starch solution for production of, e.g.

XX monosaccharides.

XX Claim 18; Page 55-58; 63pp; English.

XX This sequence is a truncated A. niger glucoamylase G1 protein. The
XX invention relates to a method of saccharifying a liquefied starch
XX solution comprising: (a) a saccharification step during which one or more
XX enzymatic saccharification stages (using a mutated Aspergillus niger
XX glucoamylase) take place; (b) one or more high temperature membrane
XX separation steps; and (c) re-circulation of the saccharification enzyme;
XX in which the membrane separation steps are carried out as an integral
XX part of the saccharification step. The method can be used for the
XX production of mono and/or oligosaccharides from starch, including
XX dextrose, trehalose, isomaltoligosaccharides, cyclodextrins or
XX maltoligosaccharides. The use of the membrane separation step improves
XX the efficiency of the process and improves yields and purity

XX Sequence 534 AA.

XX Query Match 100.0%; Score 2771; DB 2; Length 534;

XX Best Local Similarity 100.0%; Pred. No. 1.5e-214;

XX Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 MSFRSLALSLGLVCTGLANYSKRATIDSWLSNEATYARFAILLNIGADGAWVSGADSGI 60

XX 1 MSFRSLALSLGLVCTGLANYSKRATIDSWLSNEATYARFAILLNIGADGAWVSGADSGI 60

XX 61 VVASPTDNPDYFTYTWTRDSGLVKTIVDFRNGDTSLSLTENIYISAQIVOGISNPSG 120

XX 61 VVASPTDNPDYFTYTWTRDSGLVKTIVDFRNGDTSLSLTENIYISAQIVOGISNPSG 120

QY 121 DLSSGAGLCEPKFENVDETAVTGSGWGPQRDPALRATAMI GFGQWLLDNGYSTATDIYW 180

DB 121 DLSSGAGLCEPKFENVDETAVTGSGWGPQRDPALRATAMI GFGQWLLDNGYSTATDIYW 180

QY 181 PLVRNDLSVVAQYWNQGYDLMEVNGSSPFTTAVOHRALVEGSARFATVSSCSWCDSQ 240

DB 181 PLVRNDLSVVAQYWNQGYDLMEVNGSSPFTTAVOHRALVEGSARFATVSSCSWCDSQ 240

QY 241 APEILCYLOSFWTGSFTLANFDSRSRSGKANTLLGSIHTFDPPEACDSDTFQCSPRALA 300

DB 241 APEILCYLOSFWTGSFTLANFDSRSRSGKANTLLGSIHTFDPPEACDSDTFQCSPRALA 300

QY 301 NHEKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTIAAABOLYDALYOND 360

DB 301 NHEKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTIAAABOLYDALYOND 360

QY 361 KQSLLEVTVSLDFFKALYSDAATGYSSSSSTYSSIVDAVTFADPFVSIYETHAASNG 420

DB 361 KQSLLEVTVSLDFFKALYSDAATGYSSSSSTYSSIVDAVTFADPFVSIYETHAASNG 420

QY 421 SMSFQYDKSDGEQLSARDLTWSYALLTANNRRNSVVPASWGETSASVPGTCAATSALG 480

DB 421 SMSFQYDKSDGEQLSARDLTWSYALLTANNRRNSVVPASWGETSASVPGTCAATSALG 480

QY 481 TVSSVTVTSPSIVATGTTTATPTGSGSVTSTKTTATASKSTTTTRSGMSL 534

DB 481 TVSSVTVTSPSIVATGTTTATPTGSGSVTSTKTTATASKSTTTTRSGMSL 534

RESULT 2

AA18090
ID AA18090 standard; protein; 534 AA.

XX AA18090;

XX 02-SEP-1999 (first entry)

XX Aspergillus niger glucoamylase enzyme.

XX Glucoamylase enzyme; saccharification; starch hydrolyase;

XX dextrose syrup; syrup; ethanol; fuel; beverage; citric acid;

XX ascorbic acid; lysine; glutamic acid.

XX Aspergillus niger.

XX MO9928448-A1.

XX 10-JUN-1999.

XX 26-NOV-1998; 98WO-DK000520.

XX 26-NOV-1997; 97US-00979673.

XX 30-DEC-1997; 97DK-00001557.

XX 30-JUN-1998; 98US-00107657.

XX 10-JUL-1998; 98DK-00000925.

XX (NOVO) NOVO-NORDISK AS.

XX Nielsen BR, Nielsen RI, Lehmebeck J;

XX WPI; 1999-404822/34.

XX N-PSDB; AAX81797.

XX New glucoamylase obtained from *Talaromyces emersonii*.

XX Disclosure; Page 63-64; 79pp; English.

XX The present sequence represents a glucoamylase enzyme obtained from
XX Aspergillus niger. The specification describes a *Talaromyces emersonii*
XX glucoamylase enzyme. The glucoamylase enzymes have high thermal stability
XX so that a saccharification process may be carried out within a shorter
XX period of time or the process may be carried out using a lower enzyme
XX dosage. The glucoamylase enzymes can be used for saccharifying starch

CC hydrolyzate for converting starch or partially hydrolyzed starch into a
CC syrup containing dextrose. They can be used for producing
CC oligosaccharides, specialty syrups, ethanol for fuel, beverages or
CC organic compounds such as citric acid, ascorbic acid, lysine or glutamic
CC acid

SQ Sequence 534 AA;

Query Match 100.0%; Score 2771; DB 2; Length 534;

Best Local Similarity 100.0%; Pred. No. 1.5e-214; Mismatches 0; Indels 0; Gaps 0;

Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRSLALSGVCTGLANVSKRATLDMSLNEATVARTAILNNIGADGAWVSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVSKRATLDMSLNEATVARTAILNNIGADGAWVSGADSGI 60
QY 61 VVASSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYQGISNPG 120
DB 61 VVASSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYQGISNPG 120
QY 121 DLSGAGLGEPEFNVDETAYTSGWGRPQRDGPALRTAMIGQWMLDNGYSTATDIW 180
DB 121 DLSGAGLGEPEFNVDETAYTSGWGRPQRDGPALRTAMIGQWMLDNGYSTATDIW 180
QY 181 PLVRNDLSVVAQYMNQGYDLMEVNGSSFTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
DB 181 PLVRNDLSVVAQYMNQGYDLMEVNGSSFTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
QY 241 APEILICYLOSFWTGSFTILANFDSSRSRSGKANTLLGSIHFDPEAACDDSTFQPCSPRALA 300
DB 241 APEILICYLOSFWTGSFTILANFDSSRSRSGKANTLLGSIHFDPEAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGMPFLCTLAAEQYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGMPFLCTLAAEQYDALYQMD 360
QY 361 KQGSLEVTDVSLDFKALYSDAATGTYSSSSSTYSISVDAVKTFFADGFSIVETHAASNG 420
DB 361 KQGSLEVTDVSLDFKALYSDAATGTYSSSSSTYSISVDAVKTFFADGFSIVETHAASNG 420
QY 421 SMSBOYDKSDGEQLSARDLTWGYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSBOYDKSDGEQLSARDLTWGYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TYSSVTWTSWPSIVATGTTTATPTGSGSVSTSKTTATASKTSTTTSSGMSL 534
DB 481 TYSSVTWTSWPSIVATGTTTATPTGSGSVSTSKTTATASKTSTTTSSGMSL 534

RESULT 3
AAB03450
ID AAB03450 standard; protein; 534 AA.
XX
AC AAB03450;
XX
DT 03-JAN-2001 (first entry)
XX
DE Aspergillus niger G2 glucoamylase.
XX
KW Glucoamylase G2; starch hydrolysis; high fructose corn syrup;
XX thermal stability.
XX
OS Aspergillus niger.
XX
FH Key
FH Peptide 1..27
FT Cleavage-site /label= prepropeptide_sequence
FT /label= 27..28
FT /label= tripeptidyl_aminopeptidase_cleavage_site
FT Protein 28..534
XX /label= mature_glucoamylase
XX
PN WO200034452-A1.

XX 15-JUN-2000.
PD 07-DEC-1999; 99WO-DK000686.
XX
PF 07-DEC-1999; 99WO-DK000686.
XX
PR 07-DEC-1998; 98DK-00001616.
PR 24-MAR-1999; 99DK-00000409.
XX
XX
PA (NOVO) NOVO-NORDISK AS.
XX
XX Nielsen BR, Svendsen A, Boisen K, Vind J, Pedersen H;
PI WPI; 2000-431296/37.
XX
DR Variance of parent fungal glucoamylase useful for producing ethanol,
PT beverages and in fermentation processes comprise a peptide extension at
PT the amino terminal.
XX
PS Disclosure; Page 50-51; 61pp; English.
XX
CC The present sequence is the G2 glucoamylase from Aspergillus niger. It
CC was used to construct a version of the enzyme which has improved thermal
CC stability, and which can be used more efficiently in starch hydrolysis.
CC This process is used during high fructose corn syrup production, as well
CC as in the production of ethanol for fuel or beverage, and in fermentation
CC processes for producing organic compounds such as citric acid, ascorbic
CC acid, lysine and glutamic acid
XX
SQ Sequence 534 AA;

Query Match 100.0%; Score 2771; DB 3; Length 534;

Best Local Similarity 100.0%; Pred. No. 1.5e-214; Mismatches 0; Indels 0; Gaps 0;

Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRSLALSGVCTGLANVSKRATLDMSLNEATVARTAILNNIGADGAWVSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVSKRATLDMSLNEATVARTAILNNIGADGAWVSGADSGI 60
QY 61 VVASSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYQGISNPG 120
DB 61 VVASSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYQGISNPG 120
QY 121 DLSGAGLGEPEFNVDETAYTSGWGRPQRDGPALRTAMIGQWMLDNGYSTATDIW 180
DB 121 DLSGAGLGEPEFNVDETAYTSGWGRPQRDGPALRTAMIGQWMLDNGYSTATDIW 180
QY 181 PLVRNDLSVVAQYMNQGYDLMEVNGSSFTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
DB 181 PLVRNDLSVVAQYMNQGYDLMEVNGSSFTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
QY 241 APEILICYLOSFWTGSFTILANFDSSRSRSGKANTLLGSIHFDPEAACDDSTFQPCSPRALA 300
DB 241 APEILICYLOSFWTGSFTILANFDSSRSRSGKANTLLGSIHFDPEAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGMPFLCTLAAEQYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGMPFLCTLAAEQYDALYQMD 360
QY 361 KQGSLEVTDVSLDFKALYSDAATGTYSSSSSTYSISVDAVKTFFADGFSIVETHAASNG 420
DB 361 KQGSLEVTDVSLDFKALYSDAATGTYSSSSSTYSISVDAVKTFFADGFSIVETHAASNG 420
QY 421 SMSBOYDKSDGEQLSARDLTWGYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSBOYDKSDGEQLSARDLTWGYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TYSSVTWTSWPSIVATGTTTATPTGSGSVSTSKTTATASKTSTTTSSGMSL 534
DB 481 TYSSVTWTSWPSIVATGTTTATPTGSGSVSTSKTTATASKTSTTTSSGMSL 534

RESULT 4
AA77740

ID AAY77740 standard; protein; 534 AA.
 XX
 AC AAY77740;
 XX
 DT 22-MAY-2000 (first entry)
 XX
 DE A. niger G2 glucamylase.
 XX
 KW Glucoamylase; variant; starch conversion; saccharification; ethanol;
 KM fuel; beverage; fermentation; citric acid; ascorbic acid; thermostable;
 XX G2 glucamylase; fungl.
 XX
 OS Aspergillus niger.
 XX
 PN WO200004136-A1.
 XX
 PD 27-JAN-2000.
 XX
 PF 09-JUL-1999; 99MO-DK000392.
 XX
 PR 15-JUL-1998; 98DK-00000937.
 PR 17-DEC-1998; 98DK-00001667.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Nielsen BR, Svendsen A, Pedersen H, Vind J, Hendriksen HV;
 PI Frandsen TP;
 DR WPI; 2000-182412/16.
 DR N-PSDB; AAB87842.
 PT Variant fungal glucoamylases with improved thermostability and increased
 PT specific activity, useful in saccharification processes.
 XX
 PS Claim 1; Page 78-81; 116pp; English.
 XX
 CC The invention relates to variant fungal glucoamylases comprising specific
 CC mutations. The variant comprises one or more mutation in position/region
 CC 1-18, 19-35, 40-62, 73-80, 93-127, 170-184, 200-212, 234-246, 287-319,
 CC 334-341, 353-374, 388-414, 445-470 of the parent G2 glucoamylase sequence
 CC (AAY77740) from A. niger, with the exception of: N20C, A27C, S30P, Y48W,
 CC Y50F, W52F, R54K/L, D55G/V, G57A, K108R, D112Y, Y116A/W, S119C/W/E/G/Y/P,
 CC W120H/L/F/Y, G121T/A, R122Y, P123G, Q124H, R125K, W170F, N171S, Q172N, CC
 T173G, G174C, I175F, D176N/E, L177H/D, W178R/D, E179Q/D, E180D/Q,
 CC V181D/A/T, N182A/D/Q/Y/S, G183K, S184H, W212F, R241K, A246C, D293E/Q,
 CC A302Y, R305K, Y306F, D309N/E, Y312W, W317F, E389D/Q, H391W, A392D, A393P,
 CC N395Q, G396S, E400Q/C, Q401R, G407D, E408P, L410F, S411A/G/C/H/D, and
 CC S460P. The glucoamylase variants are useful in a starch conversion
 CC process, especially continuous process which include a continuous
 CC saccharification process. The variants can be used for producing
 CC oligosaccharides, specialty syrups, or ethanol for fuel or beverages.
 CC They can also be used in fermentation processes for producing organic
 CC compounds such as citric acid, ascorbic acid, lysine and glutamic acid.
 CC The glucoamylase variants have improved thermostability and/or increased
 CC specific activity. This is advantageous in industrial saccharification
 CC processes. The risk of microbial contamination is also reduced when
 CC carrying the saccharification process at temperatures above 63 plusOC. An
 CC increased specific activity towards short chain saccharides such as
 CC maltose (without reducing the activity towards oligosaccharides) would
 CC also permit using a lower enzyme dosage and/or shorter process times. The
 CC present sequence represents the G2 glucoamylase from A. niger, the parent
 CC glucoamylase used for constructing the variants
 XX
 SQ Sequence 534 AA:

Query Match 100.0%; Score 2771; DB 3; Length 534;
 Best Local Similarity 100.0%; Pred. No. 1.5e-214;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRSLALSLGLVCTGLIANYISKRAITDLSWLSNEATYATTAIINIGADGAWVSGADSGI 60
 DB 1 MSFRSLALSLGLVCTGLIANYISKRAITDLSWLSNEATYATTAIINIGADGAWVSGADSGI 60

QY 61 VVASPESTDNDPDYFYTTWTRDSGLVLTUVDLFRNGDTSLLSTIENTYISAQAIYOGISNPBG 120
 DB 61 VVASPESTDNDPDYFYTTWTRDSGLVLTUVDLFRNGDTSLLSTIENTYISAQAIYOGISNPBG 120
 QY 121 DLSSGAGLGEPKFKNVDETAVTGSGWGRPORBDGALRATAMIGFCOMLLDNGYSTATDIYW 180
 DB 121 DLSSGAGLGEPKFKNVDETAVTGSGWGRPORBDGALRATAMIGFCOMLLDNGYSTATDIYW 180
 QY 181 PLVRNDLSVYAOYMNQGYDLMEEVNGSSFFTIAYOHRALVEGSAPATVAGSSCWDSQ 240
 DB 181 PLVRNDLSVYAOYMNQGYDLMEEVNGSSFFTIAYOHRALVEGSAPATVAGSSCWDSQ 240
 QY 241 APEIILCYLGSFWTGSFTLANPDSRSRGKANTLLGSIHFDPPEAACDDSTPOCSPRALA 300
 DB 241 APEIILCYLGSFWTGSFTLANPDSRSRGKANTLLGSIHFDPPEAACDDSTPOCSPRALA 300
 QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMD 360
 DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMD 360
 QY 361 KQGSLEVTDVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTADGFSIVETHAASNG 420
 DB 361 KQGSLEVTDVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTADGFSIVETHAASNG 420
 QY 421 SMSEQYDKSDGBQLSARDLTWSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAG 480
 DB 421 SMSEQYDKSDGBQLSARDLTWSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAG 480
 QY 481 TYSSVTVTSPSPSVATGCTTTTATPTGSGSVTSTKTTATASKTSTYTSAGMSL 534
 DB 481 TYSSVTVTSPSPSVATGCTTTTATPTGSGSVTSTKTTATASKTSTYTSAGMSL 534

RESULT 5

ID AAB48171 standard; protein; 534 AA.

AC AAB48171;

DT 02-APR-2001 (first entry)

XX A. niger G1 glucoamylase polypeptide.

KW Thermosacus crustaceus; glucoamylase; starch conversion; ethanol;

KM maltose syrup; beverage; citric acid; ascorbic acid; detergent;

XX thermostability; glucose; G1 glucoamylase.

OS Aspergillus niger.

PH Key Location/Qualifiers

FT Peptide 1..24 /note="signal peptide"

FT Protein 25..534 /note="mature protein"

PN WO200075296-A1.

PD 14-DEC-2000.

PF 02-JUN-2000; 2000MO-DK000301.

PR 02-JUN-1999; 99DK-00000779.

XX (NOVO) NOVO NORDISK AS.

PI Nielsen BR, Kauppinen MS, Nielsen RI;

DR WPI; 2001-071066/08.

PT N-PSDB; AAC84444.

Novel Thermosacus crustaceus glucoamylase polypeptide useful in a continuous starch conversion process, detergents, ethanol for fuel or drinking, and oligosaccharides, specialty syrups, ethanol for fuel or drinking, and

QY 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360
 Db 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360
 QY 361 KQGSLEVTQVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420
 Db 361 KQGSLEVTQVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420
 QY 421 SMEQYDKSDGEQLSARDLTWSTYALILTANRRNSVVPASWGETSASSVPGTCAATSAIG 480
 Db 421 SMEQYDKSDGEQLSARDLTWSTYALILTANRRNSVVPASWGETSASSVPGTCAATSAIG 480
 QY 481 TYSSVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 534
 Db 481 TYSSVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 534
 RESULT 7
 AAY77741
 ID AAY77741 standard; protein; 640 AA.
 XX
 AC AAY77741;
 XX
 DT 22-MAY-2000 (first entry)
 XX
 DE A. niger G1 glucoamylase.
 XX
 KW glucoamylase; variant; starch conversion; saccharification; ethanol;
 KM fuel; beverage; fermentation; citric acid; ascorbic acid; thermostable;
 XX G1 glucoamylase; fungal.
 OS Aspergillus niger.
 PN WO200004136-A1.
 PD 27-JAN-2000.
 XX
 PF 09-JUL-1999; 99NO-DK000392.
 XX
 PR 15-JUL-1998; 98DK-00000937.
 PR 17-DEC-1998; 98DK-00001667.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Nielsen BR, Svendsen A, Pedersen H, Vind J, Hendriksen HV;
 PI Frandsen TP;
 XX
 DR N-PSDB; AAZ87843.
 DR
 XX
 PT Variant fungal glucoamylases with improved thermostability and increased
 XX specific activity, useful in saccharification processes.
 PS
 XX Disclosure; Page 91-93; 116pp; English.
 CC The invention relates to variant fungal glucoamylases. The variants
 CC comprise specific mutations in the parent G2 glucoamylase (AMG) sequence
 CC (AAY77740) from A. niger (see AAZ87842 for specific positions of the
 CC mutations). The glucoamylase variants are useful in a starch conversion
 CC process, especially continuous process which include a continuous
 CC saccharification process. The variants can be used for producing
 CC oligosaccharides, specially syrups, or ethanol for fuel or beverages.
 CC They can also be used in fermentation processes for producing organic
 CC compounds such as citric acid, ascorbic acid, lysine and glutamic acid.
 CC The glucoamylase variants have improved thermostability and/or increased
 CC specific activity. This is advantageous in industrial saccharification
 CC processes. The risk of microbial contamination is also reduced when
 CC carrying the saccharification process at temperatures above 63 plusOC. An
 CC increased specific activity towards short chain saccharides such as
 CC maltose (without reducing the activity towards oligosaccharides) would
 CC also permit using a lower enzyme dosage and/or shorter process times. The
 CC present sequence represents the G1 glucoamylase from A. niger
 ,XX

SEQ Sequence 640 AA:
 Query Match 99.0%; Score 2742; DB 3; Length 640;
 Best Local Similarity 99.6%; Pred. No. 4,4e-212;
 Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MFSRSLATSGVCTGLAVISKRATLDSWLSNEATVARTALINNGADGAWSGADSGI 60
 Db 1 MFSRSLATSGVCTGLAVISKRATLDSWLSNEATVARTALINNGADGAWSGADSGI 60
 QY 61 VVASPESTDNDPDEYFTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYOGISNPSG 120
 Db 61 VVASPESTDNDPDEYFTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYOGISNPSG 120
 QY 121 DLSSGAGLGEKFNVDETAYTSGWGPORDGAPALRTAMIGFGQWLLDNGYSTARDIYW 180
 Db 121 DLSSGAGLGEKFNVDETAYTSGWGPORDGAPALRTAMIGFGQWLLDNGYSTARDIYW 180
 QY 181 PLVRNDLSYVAQWNGOTGYDIMEVNGSFFTTAVOHRALVEGSAFATVAGSSCWCDSQ 240
 Db 181 PLVRNDLSYVAQWNGOTGYDIMEVNGSFFTTAVOHRALVEGSAFATVAGSSCWCDSQ 240
 QY 241 APEILCYLQSFMTGFTLANFDSRSRGKDANTLLGSIHTFDPACDDSTFQCSFRALA 300
 Db 241 APEILCYLQSFMTGFTLANFDSRSRGKDANTLLGSIHTFDPACDDSTFQCSFRALA 300
 QY 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360
 Db 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360
 QY 361 KQGSLEVTQVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420
 Db 361 KQGSLEVTQVSLDFPKALYSDAATGYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420
 QY 421 SMEQYDKSDGEQLSARDLTWSTYALILTANRRNSVVPASWGETSASSVPGTCAATSAIG 480
 Db 421 SMEQYDKSDGEQLSARDLTWSTYALILTANRRNSVVPASWGETSASSVPGTCAATSAIG 480
 QY 481 TYSSVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 530
 Db 481 TYSSVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 530
 RESULT 8
 AAB61905
 ID AAB61905 standard; protein; 640 AA.
 XX
 AC AAB61905;
 XX
 DT 08-MAY-2001 (first entry)
 XX
 DE A. niger protein sequence id No. 13.
 XX
 KW Glucoamylase; variant; starch; dextrose; saccharification; maltodextrin;
 KM glucose syrup; fuel; ethanol; beverage; fermentation; citric acid;
 XX ascorbic acid.
 OS Aspergillus niger.
 PN WO200104273-A2.
 PD 18-JAN-2001.
 XX
 PF 07-JUL-2000; 2000WO-DK000373.
 XX
 PR 09-JUL-1999; 99DK-00000999.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Nielsen BR, Svendsen A, Pedersen H, Vind J, Hendriksen HV;
 PI Frandsen TP;
 XX
 DR WPI; 2001-138334/14.
 ,XX

XX Novel variant of parent glucoamylase useful in starch conversion process,
 PT and for producing oligosaccharides, maltodextrins, glucose syrups, fuel,
 PT drinking ethanol, beverage and organic compounds.

XX Disclosure, Page 57-58; 58pp; English.

XX The invention relates to a variant of a parent glucoamylase, comprising
 CC an alteration at positions 59, 66, 72, 119, 189, 223, 227, 313, 340, 342,
 CC 352, 379, 386, 393, 395, 402, 408, 416, 425, 442, 444, 486, 490, 494,
 CC where the alteration is independently an insertion, substitution or
 CC deletion of amino acid which occupies the position. The variant
 CC glucoamylase is useful for converting starch or partially hydrolyzed
 CC starch into a syrup containing dextrose, by saccharifying starch
 CC hydrolyzate. The variant is useful in the starch conversion process, for
 CC producing oligosaccharides, maltodextrins or glucose syrups, fuel,
 CC drinking ethanol, beverage and in a fermentation process for producing
 CC organic compounds, such as citric acid, ascorbic acid, lysine and
 CC glutamic acid. It is useful for improving the thermal stability and/or
 CC specific activity of a parent glucoamylase

XX Sequence 640 AA:

Query Match 99.0%; Score 2742; DB 4; Length 640;
 Best Local Similarity 99.6%; Pred. No. 4.4e-212;
 Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSFRSLALSGVCTGLANVISKRATLDMSLNEAVARTAILNINIGADGAWVSGADSGI 60
 DB 1 MSFRSLALSGVCTGLANVISKRATLDMSLNEAVARTAILNINIGADGAWVSGADSGI 60
 QY 61 VVASPTDNDPDIFFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIYVGISNPSG 120
 DB 61 VVASPTDNDPDIFFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIYVGISNPSG 120
 QY 121 DLSSGAGLGEPRFNVDETAVTGSMGRPQRDGPALRAATAMIGFQWLLDNGYSTATDIYW 180
 DB 121 DLSSGAGLGEPRFNVDETAVTGSMGRPQRDGPALRAATAMIGFQWLLDNGYSTATDIYW 180
 QY 181 PLVRNDLSVVAQYMNQGYDLMEEVNNGSFFITIAVGRALVBSAATAVGSSCSWCDSQ 240
 DB 181 PLVRNDLSVVAQYMNQGYDLMEEVNNGSFFITIAVGRALVBSAATAVGSSCSWCDSQ 240
 QY 241 APEILCYLOSFWTGSFIILANFDSRSRGKANTLLGSIHFDPPEACDDSTPOCSPRALA 300
 DB 241 APEILCYLOSFWTGSFIILANFDSRSRGKANTLLGSIHFDPPEACDDSTPOCSPRALA 300
 QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQMD 360
 DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQMD 360
 QY 361 KQGSLEVDVSLDFFPALKYSDAATGTYSSTSSSTYSISYDAVKTFAAGFYSIVETHAASNG 420
 DB 361 KQGSLEVDVSLDFFPALKYSDAATGTYSSTSSSTYSISYDAVKTFAAGFYSIVETHAASNG 420
 QY 421 SMSBOYDKSDGQOLSRDLTWGSYAALLTANNRRNSVVPASMGETSASVPGTCAATSAIG 480
 DB 421 SMSBOYDKSDGQOLSRDLTWGSYAALLTANNRRNSVVPASMGETSASVPGTCAATSAIG 480
 QY 481 TVSSVTVTWSPSIVATGTTTATPTGSSSVTSTKTTATASKTSTTTS 530
 DB 481 TVSSVTVTWSPSIVATGTTTATPTGSSSVTSTKTTATASKTSTTTS 530

RESULT 9

AA040212
 ID AAP40212 standard; protein; 639 AA.

AC AAP40212;

XX 25-MAR-2003 (revised)
 DT 09-JAN-1992 (first entry)
 XX

DE Sequence encoded by A.awamori glucoamylase genomic region.

XX Starch.hydrolysis; glucose.

XX Aspergillus awamori.

XX WO8402921-A.

XX 02-AUG-1984.

XX 26-JAN-1984; 84WO-US000122.

XX 28-JAN-1983; 83US-00461920.

XX 20-DEC-1983; 83US-00563941.

XX 27-APR-1987; 87US-00047552.

XX (CETU) CETUS CORP.

XX (CETU) CETUS CORP.

XX Nunberg JH, Flatgaard JE, Innis MA, Gelfand DH, Meade JH;

XX WPI, 1984-201413/32.

XX N-PSDB; AAN40166.

XX DNA sequence coding for fungal glucoamylase protein - for expression in

XX yeast etc. for prodn. of the enzyme.

XX Example; Table I, Page 21-25; 66pp; English.

XX The inventors claim a modified DNA sequence coding for fungal
 CC glucoamylase protein or its single or multiple base substitutions
 CC deletions, insertions or inversions is new (see AAN40165). It is derived
 CC from natural, synthetic or semisynthetic sources and is capable, when
 CC correctly combined with a cleaved expression vector, of expressing a non-
 CC native protein having glucoamylase activity on transformation of a host
 CC organism by the vector. (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 639 AA:

Query Match 96.8%; Score 2683.5; DB 1; Length 639;
 Best Local Similarity 98.1%; Pred. No. 2.3e-207;
 Matches 520; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 1 MSFRSLALSGVCTGLANVISKRATLDMSLNEAVARTAILNINIGADGAWVSGADSGI 60
 DB 1 MSFRSLALSGVCTGLANVISKRATLDMSLNEAVARTAILNINIGADGAWVSGADSGI 60
 QY 61 VVASPTDNDPDIFFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIYVGISNPSG 120
 DB 61 VVASPTDNDPDIFFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIYVGISNPSG 120
 QY 121 DLSSGAGLGEPRFNVDETAVTGSMGRPQRDGPALRAATAMIGFQWLLDNGYSTATDIYW 180
 DB 121 DLSSGAGLGEPRFNVDETAVTGSMGRPQRDGPALRAATAMIGFQWLLDNGYSTATDIYW 180
 QY 181 PLVRNDLSVVAQYMNQGYDLMEEVNNGSFFITIAVGRALVBSAATAVGSSCSWCDSQ 240
 DB 181 PLVRNDLSVVAQYMNQGYDLMEEVNNGSFFITIAVGRALVBSAATAVGSSCSWCDSQ 240
 QY 241 APEILCYLOSFWTGSFIILANFDSRSRGKANTLLGSIHFDPPEACDDSTPOCSPRALA 300
 DB 241 APEILCYLOSFWTGSFIILANFDSRSRGKANTLLGSIHFDPPEACDDSTPOCSPRALA 300
 QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQMD 360
 DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQMD 360
 QY 361 KQGSLEVDVSLDFFPALKYSDAATGTYSSTSSSTYSISYDAVKTFAAGFYSIVETHAASNG 420
 DB 361 KQGSLEVDVSLDFFPALKYSDAATGTYSSTSSSTYSISYDAVKTFAAGFYSIVETHAASNG 420
 QY 421 SMSBOYDKSDGQOLSRDLTWGSYAALLTANNRRNSVVPASMGETSASVPGTCAATSAIG 480

```

Db      420  SNEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGERTSASVPGTCAATSAIG 479
Qy      481  TVSSVTVTSMPISIVATGCTTTTATPTGSGSVTSTSKTTATASXTSTTTRS 530
Db      480  TVSSVTVTSMPISIVATGCTTTTATPTGSGSVTSTSKTTATASXTSTTTRS 529

RESULT 10
AAP81876
ID      AAP81876 standard; protein; 630 AA.
XX
AC      AAP81876;
XX
DT      24-OCT-2003 (revised)
DT      25-MAR-2003 (revised)
DT      15-MAR-1992 (first entry)
XX
DE      Sequence of amyloglucosidase.
XX
KW      Enzyme; brewing; bread-making; dextrin.
XX
OS      Aspergillus niger; NCI 22343.
XX
PN      EP260160-A.
XX
PD      16-MAR-1988.
XX
PF      10-JUN-1987; 87EP-00401300.
XX
PR      10-JUN-1986; 86PR-00008387.
PR      13-APR-1987; 87PR-00005207.
PR      13-APR-1987; 87PR-00005208.
XX
PA      (TRGE ) TRANSGENE SA.
XX
PI      Labat N, Loison G, Lemoine Y;
XX
DR      WPI; 1988-072593/11.
DR      N-PSDB; AAN82019.
XX
PT      New DNA block for expressing amyloglucosidase in yeast - contg. gene plus
PT      transcription and export sequences, and transformed cells useful in
PT      brewing and bread making.
XX
PS      Example; Fig 2; 38pp; French.
XX
CC      EcoRI-Sal I fragments of Aspergillus niger NCI 22343 were cloned in
CC      pBR32 and two clones contg. portions of the gene isolated using a pool
CC      of three probes-TG282, TG283, TG284- (AAN82014-6) designed on the basis
CC      of already published SQs. The clones were designated PTG1830 and 1831. A
CC      CDNA bank of A.niger NCI 22343 was screened for amyloglucosidase using
CC      probe TG433 (AAN82017) which corresp. to AAs 163-172 of the mature
CC      protein (see AAN82019). Oligo TG387 (AAN82018) was used to confirm the
CC      identity of the clones detected. (Updated on 25-MAR-2003 to correct PR
CC      field.) (Updated on 25-MAR-2003 to correct PA field.) (Updated on 24-OCT-
CC      2003 to standardise OS field)
XX
SQ      Sequence 630 AA;
XX
Query Match      95.9%; Score 2658; DB 1; Length 630;
Best Local Similarity 97.2%; Pred. No. 2,6e-205;
Matches 515; Conservative 2; Mismatches 3; Indels 10; Gaps 1;

```

```

Db      121  DUSGAGLGEPRKFNVDETATYTGSMGRPORD-----GGQWLLDNGYSTAIDIW 170
Qy      181  PLVRNDLSVVAQYWNQGYDLMEEVNSSFPIAVQHRALVEGSAPATAVSSCSWCDQ 240
Db      171  PLVRNDLSVVAQYWNQGYDLMEEVNSSFPIAVQHRALVEGSAPATAVSSCSWCDQ 220
Qy      241  APEILICYLOSFWTGSITLANPDSRSRGKDNATLLGSIHFPDPAAACDDSTFOCSPRALA 300
Db      231  APEILICYLOSFWTGSITLANPDSRSRGKDNATLLGSIHFPDPAAACDDSTFOCSPRALA 290
Qy      301  NHEKVVDSFRSITYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABEQLDALYQMD 360
Db      291  NHEKVVDSFRSITYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABEQLDALYQMD 350
Qy      361  KQGLEVTVDVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTFFADGFVSIYETHAASNG 420
Db      351  KQGLEVTVDVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTFFADGFVSIYETHAASNG 410
Qy      421  SNEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGERTSASVPGTCAATSAIG 480
Db      411  SNEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGERTSASVPGTCAATSAIG 470
Qy      481  TVSSVTVTSMPISIVATGCTTTTATPTGSGSVTSTSKTTATASXTSTTTRS 530
Db      471  TVSSVTVTSMPISIVATGCTTTTATPTGSGSVTSTSKTTATASXTSTTTRS 520

RESULT 11
AAB15176
ID      AAB15176 standard; protein; 616 AA.
XX
AC      AAB15176;
XX
DT      11-DEC-2000 (first entry)
XX
DE      Aspergillus awamori glucamylase.
XX
KW      Glucoamylase; enzyme; carbohydrase; glucose;
KW      1,4-alpha-D-glucan glucohydrolase.
XX
OS      Aspergillus awamori.
XX
PN      WO200043504-A1.
XX
PD      27-JUL-2000.
XX
PF      10-JAN-2000; 2000MO-US000532.
XX
PR      22-JAN-1999; 99US-00236063.
XX
PA      (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
PI      Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzato R;
PI      Ford C;
XX
DR      WPI; 2000-514725/46.
XX
PT      Fungal glucamylase for selective production of glucose rather than alpha
PT      1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
PT      with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
PS      Disclosure; Page 152-153; 160pp; English.
XX
CC      The present sequence is glucamylase (1,4-alpha-D-glucan glucohydrolase;
CC      E.C. 3.2.1.3). This enzyme is a carbohydrase, and cleaves D-glucose from
CC      the nonreducing ends of maltooligosaccharides, attacking alpha-(1,4)-,
CC      and alpha-(1,6)-glucosidic bonds. The present invention relates to mutant
CC      glucamylases (see AAB15178-B15184), which have increased
CC      thermostability, increased pH optimum and reduced isomaltose formation.
CC      The mutant proteins are useful for the selective production of glucose
CC      rather than alpha-1,6 linked disaccharide isomaltose
XX

```


SQ Sequence 616 AA:

Query Match 94.8%; Score 2628; DB 3; Length 616;
Best Local Similarity 99.6%; Pred. No. 6.5e-203;
Matches 504; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
QY 25 ATLDWLSNEATVARTAIINNIIGADGAWVSGADSGIVAVSPSTNDNDPYTWRDGLV 84
DB 1 ATLDWLSNEATVARTAIINNIIGADGAWVSGADSGIVAVSPSTNDNDPYTWRDGLV 60
QY 85 KTLVDFRNGDTSLSLTSTENYISAQAIYVQISNPSGDLSSGAGLGEPKRVDETAATGSM 144
DB 61 KTLVDFRNGDTSLSLTSTENYISAQAIYVQISNPSGDLSSGAGLGEPKRVDETAATGSM 120
QY 145 GRPORDPALRATAMIGFGOMLNDNGYSTATDIWPLVRNDLSYVAQYMNQGYDLME 204
DB 121 GRPORDPALRATAMIGFGOMLNDNGYSTATDIWPLVRNDLSYVAQYMNQGYDLME 180
QY 205 VNGSSEFTTAVOHRALVEGSAFATAVSGSCWCDQAPEILCYLQSFMTGSPILANFDSS 264
DB 181 VNGSSEFTTAVOHRALVEGSAFATAVSGSCWCDQAPEILCYLQSFMTGSPILANFDSS 240
QY 265 RSGKDANTLLGSIHTFDPEAACDSTFOPCSPRALANHKEVDSFRTITLNDGLSDSEA 324
DB 241 RSGKDANTLLGSIHTFDPEAACDSTFOPCSPRALANHKEVDSFRTITLNDGLSDSEA 300
QY 325 VAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMDKQSLKVTVDVSLDFKALYSDAAT 384
DB 301 VAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMDKQSLKVTVDVSLDFKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNGSMSEQYDKSDGEOLASADLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNGSMSEQYDKSDGEOLASADLTWSYA 420
QY 445 ALLTANNRNSVVPASWGETSASSVPGTCAATSAIGTSSVTTSPMSIVATGCTTTAT 504
DB 421 ALLTANNRNSVVPASWGETSASSVPGTCAATSAIGTSSVTTSPMSIVATGCTTTAT 480
QY 505 PTGSGSVTSTKTTATASKTSTTS 530
DB 481 PTGSGSVTSTKTTATASKTSTTS 506

RESULT 12
AAW55979 standard; protein; 616 AA.
ID AAW55979 standard; protein; 616 AA.
AC AAW55979;
XX
XX 27-JUL-1998 (first entry)
XX
DE Aspergillus awamori glucoamylase mutant S411A.
KW Aspergillus awamori; glucoamylase; Aspergillus sp; mutant; fungal; food;
KW fructose; corn; sweetener; 1,4-alpha-D-glucan glucosylhydrolase;
KW genetic engineering.
XX
XX Synthetic.
OS Aspergillus awamori.
OS
XX WO9803639-A1.
XX
XX 29-JAN-1998.
XX
XX 24-JUL-1997; 97WO-US012983.
XX
XX 24-JUL-1996; 96US-0022578P.
XX
XX 02-AUG-1996; 96US-0023077P.
XX
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
XX Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
XX Ford C;
PI
```

XX WPI; 1998-120764/11.
XX
XX Genetically engineered fungal glucoamylase - useful in, e.g. food
PT industry for production of high fructose corn sweeteners.
XX
XX Claim 10; Page; 97pp; English.

PS The present sequence represents a specifically claimed mutant
XX glucoamylase from *Aspergillus awamori* (1,4-alpha-D-glucan
CC glucoamylase). The present invention describes fungal glucoamylases
CC (FG) comprising: a mutation pair Asn20Cys coupled with Ala27Cys forming a
CC disulphide bond between the 2 members of the pair; and a 311-314loop or
CC Ser113Ala mutation. FG can be used in industry for the production of high
CC fructose corn sweeteners, while the glucose produced by glucoamylase can
CC be crystallised or used in fermentation to produce organic products, e.g.
CC citric acid, ascorbic acid, lysine, glutamic acid or ethanol for
CC beverages and fuel. The mutations provide increased thermal stability,
CC reduced isomaltose formation and increased pH optimum. N.B. The present
CC sequence is not given in the specification but is derived from SEQ ID
CC NO:1 as stated in the claim

SQ Sequence 616 AA:

Query Match 94.7%; Score 2625; DB 2; Length 616;
Best Local Similarity 99.4%; Pred. No. 1.1e-202;
Matches 503; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
QY 25 ATLDWLSNEATVARTAIINNIIGADGAWVSGADSGIVAVSPSTNDNDPYTWRDGLV 84
DB 1 ATLDWLSNEATVARTAIINNIIGADGAWVSGADSGIVAVSPSTNDNDPYTWRDGLV 60
QY 85 KTLVDFRNGDTSLSLTSTENYISAQAIYVQISNPSGDLSSGAGLGEPKRVDETAATGSM 144
DB 61 KTLVDFRNGDTSLSLTSTENYISAQAIYVQISNPSGDLSSGAGLGEPKRVDETAATGSM 120
QY 145 GRPORDPALRATAMIGFGOMLNDNGYSTATDIWPLVRNDLSYVAQYMNQGYDLME 204
DB 121 GRPORDPALRATAMIGFGOMLNDNGYSTATDIWPLVRNDLSYVAQYMNQGYDLME 180
QY 205 VNGSSEFTTAVOHRALVEGSAFATAVSGSCWCDQAPEILCYLQSFMTGSPILANFDSS 264
DB 181 VNGSSEFTTAVOHRALVEGSAFATAVSGSCWCDQAPEILCYLQSFMTGSPILANFDSS 240
QY 265 RSGKDANTLLGSIHTFDPEAACDSTFOPCSPRALANHKEVDSFRTITLNDGLSDSEA 324
DB 241 RSGKDANTLLGSIHTFDPEAACDSTFOPCSPRALANHKEVDSFRTITLNDGLSDSEA 300
QY 325 VAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMDKQSLKVTVDVSLDFKALYSDAAT 384
DB 301 VAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMDKQSLKVTVDVSLDFKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNGSMSEQYDKSDGEOLASADLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNGSMSEQYDKSDGEOLASADLTWSYA 420
QY 445 ALLTANNRNSVVPASWGETSASSVPGTCAATSAIGTSSVTTSPMSIVATGCTTTAT 504
DB 421 ALLTANNRNSVVPASWGETSASSVPGTCAATSAIGTSSVTTSPMSIVATGCTTTAT 480
QY 505 PTGSGSVTSTKTTATASKTSTTS 530
DB 481 PTGSGSVTSTKTTATASKTSTTS 506

RESULT 13
AAB15180 standard; protein; 616 AA.
ID AAB15180 standard; protein; 616 AA.
AC AAB15180;
XX
XX 11-DEC-2000 (first entry)
XX
```

DE Aspergillus awamori mutant glucoamylase S411A substitution.
XX
KW Glucoamylase; enzyme; carbohydrase; glucose;
KM 1,4-alpha-D-glucan glucohydrolase; mutain; mutation.
XX
OS Aspergillus awamori.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 411 /note="Wild-type Ser substituted by Ala"
FT
XX
PN W0200043504-A1.
PD 27-JUL-2000.
XX
PF 10-JAN-2000; 2000WO-US000532.
XX
PR 22-JAN-1999; 99US-00236063.
XX
PA (IOWA) UNIV IOWA STATE RES FOUND INC.
XX
PI Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
PI Ford C;
XX
DR WPI; 2000-514725/46.
XX
PT Fungal glucoamylase for selective production of glucose rather than alpha
PT -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
PT with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
PS Claim 10; Page; 160pp; English.
XX
CC Glucoamylase (1,4-alpha-D-glucan glucohydrolase; E.C. 3.2.1.3) is a
CC carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of
CC maltooligosaccharides, attacking alpha-(1,4) - and alpha-(1,6)-glucosidic
CC bonds. The present invention relates to mutant glucoamylases, which have
CC increased thermostability, increased pH optimum and reduced isomaltose
CC formation. The mutant proteins are useful for the selective production of
CC glucose rather than alpha-1,6 linked disaccharide isomaltose. The present
CC sequence is one such mutant enzyme. Note: The present sequence is not
CC shown in the specification but is derived from the Aspergillus awamori
CC wild-type glucoamylase sequence given in pages 152-153 of the sequence
CC listing (SEQ ID 1)
XX
SQ Sequence 616 AA;
Query Match 94.7%; Score 2625; DB 3; Length 616;
Best Local Similarity 99.4%; Pred. NO. 1.1e-202;
Matches 503; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 25 ATLDWSLNEATVARTAILNNGADGAMVSGADSGIVVASPSTNDPDPYFTWTRDSGLVL 84
DB 1 ATLDWSLNEATVARTAILNNGADGAMVSGADSGIVVASPSTNDPDPYFTWTRDSGLVL 60
QY 85 KTLVDLFRNGDTSLSLTENYISAQAIIVGQISNPSGDLSSGAGLGEPEKFNVDETAYTGSW 144
DB 61 KTLVDLFRNGDTSLSLTENYISAQAIIVGQISNPSGDLSSGAGLGEPEKFNVDETAYTGSW 120
QY 145 GRPORDSPALRATMTAGFGQWLDNGYSTRATDVIWLVNNDLSYVAQVNNQGYDLME 204
DB 121 GRPORDSPALRATMTAGFGQWLDNGYSTRATDVIWLVNNDLSYVAQVNNQGYDLME 180
QY 205 VNGSSPFTIAVQHRALVEGSAFATAVGSSCMCDSPAILICYLQSFMTGSFILANPDS 264
DB 181 VNGSSPFTIAVQHRALVEGSAFATAVGSSCMCDSPAILICYLQSFMTGSFILANPDS 240
QY 265 RSGKDANTLLGSIHTPDEAACDDSTFQPCSPRALANHKEVDSFRSIYTLNDGLSDSEA 324
DB 241 RSGKDANTLLGSIHTPDEAACDDSTFQPCSPRALANHKEVDSFRSIYTLNDGLSDSEA 300
QY 325 VAVGRYPEDTYYNGNPFELCTLAABQLYDALYQMDKQSLSEVTDVSLDFPKALYSDAAT 384

DB 301 VAVGRYPEDTYYNGNPFELCTLAABQLYDALYQMDKQSLSEVTDVSLDFPKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEOLASADLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEOLASADLTWSYA 420
QY 445 ALLTANRRNSVVPASWGETSASVPGTCAATSAIGTSSVYTSWPSIYATGTTTAT 504
DB 421 ALLTANRRNSVVPASWGETSASVPGTCAATSAIGTSSVYTSWPSIYATGTTTAT 480
QY 505 PTGSGSVTSTSKTATASKSTTTTS 530
DB 481 PTGSGSVTSTSKTATASKSTTTTS 506
RESULT 14
ID AAM55976
XX AAM55976 standard; protein; 616 AA.
XX
AC AAM55976;
XX
DT 27-JUL-1998 (first entry)
XX
DE Aspergillus awamori glucoamylase mutant N20C, A27C.
XX
KW Aspergillus awamori; glucoamylase; Aspergillus sp; mutant; fungal; food;
KW fructose; corn; sweetener; 1,4-alpha-D-glucan glucohydrolase;
KM genetic engineering.
XX
OS Synthetic.
OS Aspergillus awamori.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 20..27
XX
PN W09803639-A1.
XX
PD 29-JAN-1998.
XX
PF 24-JUL-1997; 97WO-US012983.
XX
PR 24-JUL-1996; 96US-0022578P.
PR 02-AUG-1996; 96US-0023077P.
XX
PA (IOWA) UNIV IOWA STATE RES FOUND INC.
XX
PI Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
PI Ford C;
XX
DR WPI; 1998-120764/11.
XX
PT Genetically engineered fungal glucoamylase - useful in, e.g. food
PT industry for production of high fructose corn sweeteners.
XX
PS Claim 1; Page; 97pp; English.
XX
CC The present sequence represents a specifically claimed mutant
CC glucoamylase from Aspergillus awamori (1,4-alpha-D-glucan
CC glucosylase). The present invention describes fungal glucoamylases
CC (CG) comprising: a mutation pair Asn20Cys coupled with Ala27Cys forming a
CC disulfide bond between the 2 members of the pair; and a 311-314loop or
CC Ser411Ala mutation. PG can be used in industry for the production of high
CC fructose corn sweeteners, while the glucose produced by glucoamylase can
CC be crystallised or used in fermentation to produce organic products, e.g.
CC citric acid, ascorbic acid, lysine, glutamic acid or ethanol for
CC beverages and fuel. The mutations provide increased thermal stability,
CC reduced isomaltose formation and increased pH optimum. N.B. The present
CC sequence is not given in the specification but is derived from SEQ ID
CC NO:1 as stated in the Claim
XX
SQ Sequence 616 AA;
Query Match 94.4%; Score 2615; DB 2; Length 616;

Best Local Similarity 99.2%; Pred. No. 7.3e-202;
Matches 502; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 25 ATLDSWLSNEATVARTAIINNIIGADGAWVSGADSGIVASPSTNDNDPYFTWTRDSGLVL 84
DB 1 ATLDSWLSNEATVARTAIINNIIGADGAWVSGADSGIVASPSTNDNDPYFTWTRDSGLVL 60
QY 85 KTLVDLFRNGDTSLSLTIENYISAQAIYVGGISNPSGDLSSGAGLGPKRVNDETAATGSM 144
DB 61 KTLVDLFRNGDTSLSLTIENYISAQAIYVGGISNPSGDLSSGAGLGPKRVNDETAATGSM 120
QY 145 GRPORDGPALRATAMIGFGOMLLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 204
DB 121 GRPORDGPALRATAMIGFGOMLLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 180
QY 205 VNGSSFFTIAVOHRAIVEGSAFATAVGSSCSMCDGAPILLCYLGSPFWGSPFLANFDS 264
DB 181 VNGSSFFTIAVOHRAIVEGSAFATAVGSSCSMCDGAPILLCYLGSPFWGSPFLANFDS 240
QY 265 RSGKDANTLLGSIHTFDPPEACDDSTFQPCSPRALANHKEVDSFRSITTLNDGLSDSEA 324
DB 241 RSGKDANTLLGSIHTFDPPEACDDSTFQPCSPRALANHKEVDSFRSITTLNDGLSDSEA 300
QY 325 VAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQMDKQSLFVTDVSLDFPKALYSDAAT 384
DB 301 VAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQMDKQSLFVTDVSLDFPKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNSMSGEQYDKSGEQLSARDLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNSMSGEQYDKSGEQLSARDLTWSYA 420
QY 445 ALLTANRRNSVVPASWGETSASVPGTCAATSAIGTYSVTWTSMPISIVATGCTTTTAT 504
DB 421 ALLTANRRNSVVPASWGETSASVPGTCAATSAIGTYSVTWTSMPISIVATGCTTTTAT 480
QY 505 PTGSGSVTSTSKTTATASKTSTTRS 530
DB 481 PTGSGSVTSTSKTTATASKTSTTRS 506
```

RESULT 15

AAB15178
ID AAB15178 standard; protein; 616 AA.

AC AAB15178;

XX 11-DEC-2000 (first entry)

DE Aspergillus awamori mutant glucoamylase N20C/A27C substitution.

KW Glucoamylase; enzyme; carbohydrate; glucose;
1,4-alpha-D-glucan glucohydrolase; mucin; mutation.

XX Aspergillus awamori.

OS Synthetic.

XX Key Location/Qualifiers

FT Disulfide-bond 20..27

FT Misc-difference 20 /note= "Wild-type Asn substituted by Cys"

FT Misc-difference 27 /note= "Wild-type Ala substituted by Cys"

PN WO200043504-A1.

PD 27-JUL-2000.

PF 10-JAN-2000; 2000MO-US000532.

PR 22-JAN-1999; 99US-00236063.

XX (IOWA) UNIV IOWA STATE RES FOUND INC.

PI Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzalko R;
PI Ford C;
XX MPI; 2000-514725/46.

PT Fungal glucoamylase for selective production of glucose rather than alpha
PT -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
PT with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
PS Claim 1; Page; 160pp; English.

XX Glucoamylase (1,4-alpha-D-glucan glucohydrolase; E.C. 3.2.1.3) is a
CC carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of
CC maltotriogalactarides, attacking alpha-(1,4)-, and alpha-(1,6)-glucosidic
CC bonds. The present invention relates to mutant glucoamylases, which have
CC increased thermostability, increased pH optimum and reduced isomaltose
CC formation. The mutant proteins are useful for the selective production of
CC glucose rather than alpha-1,6 linked disaccharide isomaltose. The present
CC sequence is one such mutant enzyme. Note: The present sequence is not
CC shown in the specification but is derived from the Aspergillus awamori
CC wild-type glucoamylase sequence given in pages 152-153 of the sequence
CC listing (SEQ ID 1)

XX Sequence 616 AA;

Query Match 94.4%; Score 2615; DB 3; Length 616;

Best Local Similarity 99.2%; Pred. No. 7.3e-202;
Matches 502; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 25 ATLDSWLSNEATVARTAIINNIIGADGAWVSGADSGIVASPSTNDNDPYFTWTRDSGLVL 84
DB 1 ATLDSWLSNEATVARTAIINNIIGADGAWVSGADSGIVASPSTNDNDPYFTWTRDSGLVL 60
QY 85 KTLVDLFRNGDTSLSLTIENYISAQAIYVGGISNPSGDLSSGAGLGPKRVNDETAATGSM 144
DB 61 KTLVDLFRNGDTSLSLTIENYISAQAIYVGGISNPSGDLSSGAGLGPKRVNDETAATGSM 120
QY 145 GRPORDGPALRATAMIGFGOMLLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 204
DB 121 GRPORDGPALRATAMIGFGOMLLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 180
QY 205 VNGSSFFTIAVOHRAIVEGSAFATAVGSSCSMCDGAPILLCYLGSPFWGSPFLANFDS 264
DB 181 VNGSSFFTIAVOHRAIVEGSAFATAVGSSCSMCDGAPILLCYLGSPFWGSPFLANFDS 240
QY 265 RSGKDANTLLGSIHTFDPPEACDDSTFQPCSPRALANHKEVDSFRSITTLNDGLSDSEA 324
DB 241 RSGKDANTLLGSIHTFDPPEACDDSTFQPCSPRALANHKEVDSFRSITTLNDGLSDSEA 300
QY 325 VAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQMDKQSLFVTDVSLDFPKALYSDAAT 384
DB 301 VAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQMDKQSLFVTDVSLDFPKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNSMSGEQYDKSGEQLSARDLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNSMSGEQYDKSGEQLSARDLTWSYA 420
QY 445 ALLTANRRNSVVPASWGETSASVPGTCAATSAIGTYSVTWTSMPISIVATGCTTTTAT 504
DB 421 ALLTANRRNSVVPASWGETSASVPGTCAATSAIGTYSVTWTSMPISIVATGCTTTTAT 480
QY 505 PTGSGSVTSTSKTTATASKTSTTRS 530
DB 481 PTGSGSVTSTSKTTATASKTSTTRS 506
```

RESULT 16

AAB15184
ID AAB15184 standard; protein; 616 AA.

AC AAB15184;

DT 11-DEC-2000 (first entry)

XX A.awamori mutant glucoamylase S30P/E137A/S411A loop substitution.
DE
XX Glucoamylase; enzyme; carbohydrase; glucose;
KM 1,4-alpha-D-glucan glucosylhydrolase; mutuin; mutation.
XX
OS Aspergillus awamori.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 30 /note= "Wild-type Ser substituted by Pro"
FT Misc-difference 137 /note= "Wild-type Glu substituted by Ala"
FT /note= "Wild-type Ser substituted by Ala"
FT Misc-difference 411 /note= "Wild-type Ser substituted by Ala"
XX
PN WC0200043504-A1.
XX
PD 27-JUL-2000.
XX
PF 10-JAN-2000; 2000MO-US000532.
XX
PR 22-JAN-1999; 99US-00236063.
XX
PA (IOWA) UNIV IOWA STATE RES FOUND INC.
XX
PI Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
PI Ford C;
XX
DR WPI; 2000-514725/46.
XX
PT Fungal glucoamylase for selective production of glucose rather than alpha
PT -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
PT with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
PS Claim 45; Page; 160pp; English.
XX
XX Glucoamylase (1,4-alpha-D-glucan glucosylhydrolase; E.C. 3.2.1.3) is a
CC carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of
CC maltooligosaccharides, attacking alpha-(1,4)-, and alpha-(1,6)-glucosidic
CC bonds. The present invention relates to mutant glucoamylases, which have
CC increased thermostability, increased pH optimum and reduced isomaltose
CC formation. The mutant proteins are useful for the selective production of
CC glucose rather than alpha-1,6 linked disaccharide isomaltose. The present
CC sequence is one such mutant enzyme. Note: The present sequence is not
CC shown in the specification but is derived from the Aspergillus awamori
CC wild-type glucoamylase sequence given in pages 152-153 of the sequence
CC listing (SEQ ID 1)
XX
SQ Sequence 616 AA;
Query Match 94.3%; Score 2614; DB 3; Length 616;
Best Local Similarity 99.0%; Pred. No. 8.8e-202;
Matches 501; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 25 ATIDSWLSNEATVARTAIINNIGADGAMVSGADSGIVVASPSTDPDFFYTWTRDSGLVL 84
DB 1 ATIDSWLSNEATVARTAIINNIGADGAMVSGADSGIVVASPSTDPDFFYTWTRDSGLVL 60
QY 85 KTIIVDLFRNGDTSLISTENYISAOAIVOGISNPSGLSGAGIGEKFNVDERTAYYGSW 144
DB 61 KTIIVDLFRNGDTSLISTENYISAOAIVOGISNPSGLSGAGIGEKFNVDERTAYYGSW 120
QY 145 GRPORDPALRATMIGFGOWLDNGYTSATDITVWPLVNRNDLSYVAQYNNNGGYDLME 204
DB 121 GRPORDPALRATMIGFGOWLDNGYTSATDITVWPLVNRNDLSYVAQYNNNGGYDLME 180
QY 205 VNGSSFFTIIVQHRALVEGSAFATAVSSGSCWCDQAPELLICYQSFWTGSFILANPDS 264
DB 181 VNGSSFFTIIVQHRALVEGSAFATAVSSGSCWCDQAPELLICYQSFWTGSFILANPDS 240
QY 265 RSGKDANTLIGSIHTPEPAACDSTFQPCSPRALANHKEVVDSPRSIYTLNDGLSDSEA 324

DB 241 RSGKDANTLIGSIHTPEPAACDSTFQPCSPRALANHKEVVDSPRSIYTLNDGLSDSEA 300
QY 325 VAVGRYPEDTYNGNFWPLCTTLAAEQLYDALYQMDKQSLLEVTVSLDFPKALYSDAAT 384
DB 301 VAVGRYPEDTYNGNFWPLCTTLAAEQLYDALYQMDKQSLLEVTVSLDFPKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNSMSGEQYDKSGEQLAARDLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNSMSGEQYDKSGEQLAARDLTWSYA 420
QY 445 ALLTANNRRNSVVPASMGETSASVPGCAATSIGTSSVYVTSMPSTVATGTTTTAT 504
DB 421 ALLTANNRRNSVVPASMGETSASVPGCAATSIGTSSVYVTSMPSTVATGTTTTAT 480
QY 505 PTGSGSVTSTSKTTATATSKTSTTSS 530
DB 481 PTGSGSVTSTSKTTATATSKTSTTSS 506
RESULT 17
ID AAM55980 standard; protein; 616 AA.
XX
AC AAM55980;
XX
DT 27-JUL-1998 (first entry)
XX
DE Aspergillus awamori glucoamylase mutant N20C, A27C, S411A.
XX
KM Aspergillus awamori; glucoamylase; Aspergillus sp; mutant; fungal; food;
KM fructose; corn; sweetener; 1,4-alpha-D-glucan glucosylhydrolase;
XX genetic engineering.
OS Synthetic.
OS Aspergillus awamori.
XX
XX Key Location/Qualifiers
FH Disulfide-bond 20..27
XX
PN WC09803639-A1.
XX
PD 29-JAN-1998.
XX
PF 24-JUL-1997; 97MO-US012983.
XX
PR 24-JUL-1996; 96US-0022578P.
PR 02-AUG-1996; 96US-0023077P.
XX
PA (IOWA) UNIV IOWA STATE RES FOUND INC.
XX
PI Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
PI Ford C;
XX
DR WPI; 1998-120764/11.
XX
PT Genetically engineered fungal glucoamylase - useful in, e.g. food
PT industry for production of high fructose corn sweeteners.
XX
PS Claim 14; Page; 97pp; English.
XX
XX The present sequence represents a specifically claimed mutant
CC glucoamylase from Aspergillus awamori (1,4-alpha-D-glucan
CC glucosylhydrolase). The present invention describes fungal glucoamylases
CC (FG) comprising: a mutation pair Asn20Cys coupled with Ala27Cys forming a
CC disulfide bond between the 2 members of the pair; and a 311-314loop or
CC Ser411Ala mutation. FG can be used in industry for the production of high
CC fructose corn sweeteners, while the glucose produced by glucoamylase can
CC be crystallised or used in fermentation to produce organic products, e.g.
CC citric acid, ascorbic acid, lysine, glutamic acid or ethanol for
CC beverages and fuel. The mutations provide increased thermal stability,
CC reduced isomaltose formation and increased pH optimum. N.B. The present
CC sequence is not given in the specification but is derived from SEQ ID

CC NO.1 as stated in the claim
XX
SQ Sequence 616 AA;

Query Match 94.3%; Score 2612; DB 2; Length 616;
Best Local Similarity 99.0%; Pred. No. 1.3e-201;
Matches 501; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

25 ATLDLSLSEATVARTAILNINIGADGAWVSGADSGIVVSPSTNDPDYFTWTRDSGLV 84
1 ATLDLSLSEATVARTAILNINIGADGAWVSGADSGIVVSPSTNDPDYFTWTRDSGLV 60
85 KTLVDLFRNGDTSLSLTSTENIYSAQAIYVQISNPSGDLSSGAGLGPKNVDETAYTGSW 144
61 KTLVDLFRNGDTSLSLTSTENIYSAQAIYVQISNPSGDLSSGAGLGPKNVDETAYTGSW 120
145 GRPORDPALRATAMIGFCQWMLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 204
121 GRPORDPALRATAMIGFCQWMLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 180
205 VNGSSPFTIAYOHRALVEGSAFATAVGSSCSMCDQAPILCYLQSFMTGSFTLANFDS 264
181 VNGSSPFTIAYOHRALVEGSAFATAVGSSCSMCDQAPILCYLQSFMTGSFTLANFDS 240
265 RSGKDANTLLGSIHFPDPAACDSTFQCSPPALANKEVDSFRSIYTLNDGLSDSEA 324
241 RSGKDANTLLGSIHFPDPAACDSTFQCSPPALANKEVDSFRSIYTLNDGLSDSEA 300
325 VAVGRYPEDTYNGNPFLLCTLAAEQYDALYQMDKQSLVETDVSLEDFKALYSDAAT 384
301 VAVGRYPEDTYNGNPFLLCTLAAEQYDALYQMDKQSLVETDVSLEDFKALYSDAAT 360
385 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEOLAAADLTWSYA 444
361 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEOLAAADLTWSYA 420
445 ALLTANRRNSVVPASWGETSASSVGTCAATSAIGTSSVTVTSWPSIVATGCTTTAT 504
421 ALLTANRRNSVVPASWGETSASSVGTCAATSAIGTSSVTVTSWPSIVATGCTTTAT 480
505 PTGSGSVTSTSKTTATASKTSTTRS 530
481 PTGSGSVTSTSKTTATASKTSTTRS 506

RESULT 18
AAB15181
ID AAB15181 standard; protein; 616 AA.
XX
XX AAB15181;
XX
XX 11-DEC-2000 (first entry)
XX
XX Aspergillus awamori mutant glucamylase N20C/A27C/S411A substitution.
XX
XX Glucoamylase; enzyme; carbohydrase; glucose;
XX
XX 1,4-alpha-D-glucan glucosyltransferase; mutagen; mutation.
XX
XX Aspergillus awamori.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX
XX Disulfide-bond 20. 27
XX
XX Misc-difference 20 /note= "Wild-type Asn substituted by Cys"
XX
XX Misc-difference 27 /note= "Wild-type Ala substituted by Cys"
XX
XX Misc-difference 411 /note= "Wild-type Ser substituted by Ala"
XX
XX WO200043504-A1.
XX
XX PD 27-JUL-2000.

XX 10-JAN-2000; 2000WO-US000532.
XX
XX 22-JAN-1999; 99US-00236063.
XX
XX (IOWA) UNIV IOWA STATE RES FOUND INC.
XX
XX Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
XX Ford C;
XX
XX MPI; 2000-514725/46.

Fungal glucamylase for selective production of glucose rather than alpha-1,6 linked disaccharide isomaltose, has mutation pair Asp20Cys coupled with Ala27Cys forming disulfide bond between the two stabilizing members.

Claim 14; Page; 160p; English.

Glucamylase (1,4-alpha-D-glucan glucosyltransferase; E.C. 3.2.1.3) is a carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of maltooligosaccharides, attacking alpha-(1,4)-, and alpha-(1,6)-glucosidic bonds. The present invention relates to mutant glucamylases, which have increased thermostability, increased pH optimum and reduced isomaltose formation. The mutant proteins are useful for the selective production of glucose rather than alpha-1,6 linked disaccharide isomaltose. The present sequence is one such mutant enzyme. Note: The present sequence is not shown in the specification but is derived from the Aspergillus awamori wild-type glucamylase sequence given in pages 152-153 of the sequence listing (Seq ID 1)

CC
XX
SQ Sequence 616 AA;

Query Match 94.3%; Score 2612; DB 3; Length 616;
Best Local Similarity 99.0%; Pred. No. 1.3e-201;
Matches 501; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

25 ATLDLSLSEATVARTAILNINIGADGAWVSGADSGIVVSPSTNDPDYFTWTRDSGLV 84
1 ATLDLSLSEATVARTAILNINIGADGAWVSGADSGIVVSPSTNDPDYFTWTRDSGLV 60
85 KTLVDLFRNGDTSLSLTSTENIYSAQAIYVQISNPSGDLSSGAGLGPKNVDETAYTGSW 144
61 KTLVDLFRNGDTSLSLTSTENIYSAQAIYVQISNPSGDLSSGAGLGPKNVDETAYTGSW 120
145 GRPORDPALRATAMIGFCQWMLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 204
121 GRPORDPALRATAMIGFCQWMLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 180
205 VNGSSPFTIAYOHRALVEGSAFATAVGSSCSMCDQAPILCYLQSFMTGSFTLANFDS 264
181 VNGSSPFTIAYOHRALVEGSAFATAVGSSCSMCDQAPILCYLQSFMTGSFTLANFDS 240
265 RSGKDANTLLGSIHFPDPAACDSTFQCSPPALANKEVDSFRSIYTLNDGLSDSEA 324
241 RSGKDANTLLGSIHFPDPAACDSTFQCSPPALANKEVDSFRSIYTLNDGLSDSEA 300
325 VAVGRYPEDTYNGNPFLLCTLAAEQYDALYQMDKQSLVETDVSLEDFKALYSDAAT 384
301 VAVGRYPEDTYNGNPFLLCTLAAEQYDALYQMDKQSLVETDVSLEDFKALYSDAAT 360
385 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEOLAAADLTWSYA 444
361 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEOLAAADLTWSYA 420
445 ALLTANRRNSVVPASWGETSASSVGTCAATSAIGTSSVTVTSWPSIVATGCTTTAT 504
421 ALLTANRRNSVVPASWGETSASSVGTCAATSAIGTSSVTVTSWPSIVATGCTTTAT 480
505 PTGSGSVTSTSKTTATASKTSTTRS 530
481 PTGSGSVTSTSKTTATASKTSTTRS 506

```
RESULT 19
ID AAM55977 standard; protein; 616 AA.
XX
AC AAM55977;
XX
DT 27-JUL-1998 (first entry)
XX
DE Aspergillus awamori glucoamylase mutant N20C, A27C, S30P, G137A.
XX
KM Aspergillus awamori; glucoamylase; Aspergillus sp; mutant; fungal; food;
KM fructose; corn; sweetener; 1,4-alpha-D-glucan glucosyltransferase;
KM genetic engineering.
XX
OS Synthetic.
XX Aspergillus awamori.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 20..27
XX
PN WO9803639-A1.
XX
PD 29-JAN-1998.
XX
PF 24-JUL-1997; 97NO-US012983.
XX
PR 24-JUL-1996; 96US-0022578P.
PR 02-AUG-1996; 96US-0023077P.
XX
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
PI Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
PI Ford C;
XX
DR WPI; 1998-120764/11.
XX
PT Genetically engineered fungal glucoamylase - useful in, e.g. food
PT industry for production of high fructose corn sweeteners.
XX
PS Claim 4; Page; 97P; English.
XX
XX The present sequence represents a specifically claimed mutant
CC glucoamylase from Aspergillus awamori (1,4-alpha-D-glucan
CC glucosyltransferase). The present invention describes fungal glucoamylases
CC (FG) comprising: a mutation pair Asn20Cys coupled with Ala27Cys forming a
CC disulphide bond between the 2 members of the pair; and a 311-314loop or
CC Ser411Ala mutation. FG can be used in industry for the production of high
CC fructose corn sweeteners, while the glucose produced by glucoamylase can
CC be crystallised or used in fermentation to produce organic products, e.g.
CC citric acid, ascorbic acid, lysine, glutamic acid or ethanol for
CC beverages and fuel. The mutations provide increased thermal stability,
CC reduced isomaltose formation and increased pH optimum. N.B. The present
CC sequence is not given in the specification but is derived from SEQ ID
CC NO:1 as stated in the claim
XX
XX Sequence 616 AA;
SO
Query Match 94.0%; Score 2604; DB 2; Length 616;
Best Local Similarity 98.8%; Pred. No. 5.6e-201;
Matches 500; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 25 ATLDWSLSEATVARTAIINNGADGAVSGADSGIVVSPSTNDPDYFTTWRDGLVL 84
DB 1 ATLDWSLSEATVARTAIINNGADGAVSGADSGIVVSPSTNDPDYFTTWRDGLVL 60
QY KTLVDFLRNGDTSLSTLENTYISAOAVOGISNPSGDLSSGAGIGEPKFNVDATAYGWS 144
DB 61 KTLVDFLRNGDTSLSTLENTYISAOAVOGISNPSGDLSSGAGIGEPKFNVDATAYGWS 120
QY 145 GRPQRPDPALRATAMIGFGOMLDNGYTSATDITVWPLVLRNDLSYVAQYNNQGYDLME 204
DB 121 GRPQRPDPALRATAMIGFGOMLDNGYTSATDITVWPLVLRNDLSYVAQYNNQGYDLME 180
```

```
QY 205 VNGSFFTTIAVOHRAIVEGSAFATAVAGSSCSWCDQAPILCYLQSPMTGSFIANPDS 264
DB 181 VNGSFFTTIAVOHRAIVEGSAFATAVAGSSCSWCDQAPILCYLQSPMTGSFIANPDS 240
QY 265 RSGKDANTLLGSITHTPEPAACDSTFOPCSPRALANHEVVDSPFSIYTIANDGLSDSA 324
DB 241 RSGKDANTLLGSITHTPEPAACDSTFOPCSPRALANHEVVDSPFSIYTIANDGLSDSA 300
QY 325 VAVGRYPEDTYNGNPMFLCTIAAEQLYDALYQMDKQSLSEVTVSLDPFKALYSDAAT 384
DB 301 VAVGRYPEDTYNGNPMFLCTIAAEQLYDALYQMDKQSLSEVTVSLDPFKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNSMSBOYDKSDGEOLASADLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFSIVETHAASNSMSBOYDKSDGEOLASADLTWSYA 420
QY 445 ALLTANNRRNSVVPAMGETSASSVPGTCAATSAITGYSVTVTSPSIVATGCTTTTAT 504
DB 421 ALLTANNRRNSVVPAMGETSASSVPGTCAATSAITGYSVTVTSPSIVATGCTTTTAT 480
QY 505 PTGSGSVTSTSKTTATPASKTSTTRS 530
DB 481 PTGSGSVTSTSKTTATPASKTSTTRS 506
RESULT 20
AAB15179
ID AAB15179 standard; protein; 616 AA.
XX
AC AAB15179;
XX
DT 11-DEC-2000 (first entry)
XX
DE A. awamori mutant glucoamylase N20C/A27C/S30P/G137A substitution.
XX
KM Glucoamylase; enzyme; carbohydrase; glucose;
KM 1,4-alpha-D-glucan glucosyltransferase; mutain; mutation.
XX
OS Aspergillus awamori.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers.
FH Disulfide-bond 20..27
FT Misc-difference 20
FT /note= "Wild-type Asn substituted by Cys"
FT Misc-difference 27
FT /note= "Wild-type Ala substituted by Cys"
FT Misc-difference 30
FT /note= "Wild-type Ser substituted by Pro"
FT Misc-difference 137
FT /note= "Wild-type Gly substituted by Ala"
PN WO200043504-A1.
XX
XX 27-JUL-2000.
XX
XX 10-JAN-2000; 2000WO-US000532.
XX
XX 22-JAN-1999; 99US-00236063.
XX
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
PI Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
PI Ford C;
XX
DR WPI; 2000-514725/46.
XX
XX Fungal glucoamylase for selective production of glucose rather than alpha
XX -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
XX with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
XX Claim 4; Page; 160pp; English.
XX
```

CC Glucoamylase (1,4-alpha-D-glucan glucosyltransferase; E.C. 3.2.1.3) is a
 CC carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of
 CC maltooligosaccharides, attacking alpha-(1,4)- and alpha-(1,6)-glucosidic
 CC bonds. The present invention relates to mutant glucoamylases, which have
 CC increased thermostability, increased pH optimum and reduced isomaltose
 CC formation. The mutant proteins are useful for the selective production of
 CC glucose rather than alpha-1,6 linked disaccharide isomaltose. The present
 CC sequence is one such mutant enzyme. Note: The present sequence is not
 CC shown in the specification but is derived from the Aspergillus awamori
 CC wild-type glucoamylase sequence given in pages 152-153 of the sequence
 CC listing (SEQ ID 1)

XX Sequence 616 AA:

Query Match 94.0%; Score 2604; DB 3; Length 616;
 Best Local Similarity 98.8%; Pred. No. 5.6e-201;
 Matches 500; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 25 ATLDLWLSNEATVARTALINIGADGAWWSGADSGIVVSPSTNDPDYFTWTRDSGLVL 84
 DB 1 ATLDLWLSNEATVARTALINIGADGAWWSGADSGIVVSPSTNDPDYFTWTRDSGLVL 60
 QY 85 KTLVDLFRNGDTSLSLTIENTYISAQAIYVQGISNPSGDLSSGAGLGEKFNVDETAYTGSW 144
 DB 61 KTLVDLFRNGDTSLSLTIENTYISAQAIYVQGISNPSGDLSSGAGLGEKFNVDETAYTGSW 120
 QY 145 GRPORDPALRATAMIGFGQWLLDNGYTSATDIWPLVRNDLSYVAQYWNQGYDLMEW 204
 DB 121 GRPORDPALRATAMIGFGQWLLDNGYTSATDIWPLVRNDLSYVAQYWNQGYDLMEW 180
 QY 205 VNGSSFFTTIAVQHRALVEGSAFATAVAGSSCSWCDQAPILCYLQSFMTGSPFLANFDS 264
 DB 181 VNGSSFFTTIAVQHRALVEGSAFATAVAGSSCSWCDQAPILCYLQSFMTGSPFLANFDS 240
 QY 265 RSGKDNATLLGSIHTPDPEACDSTFQPCSPALANHEKVDSPRSITTLNDGLSDSEA 324
 DB 241 RSGKDNATLLGSIHTPDPEACDSTFQPCSPALANHEKVDSPRSITTLNDGLSDSEA 300
 QY 325 VAVGRYPEDTYVNGNPMFLCTLAABQLYDALYQMDKQSLVETDVSLEDFKALYSDAAT 384
 DB 301 VAVGRYPEDTYVNGNPMFLCTLAABQLYDALYQMDKQSLVETDVSLEDFKALYSDAAT 360
 QY 385 GTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNGSMSEQYDKSGEQLSARDLTMSYA 444
 DB 361 GTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNGSMSEQYDKSGEQLSARDLTMSYA 420
 QY 445 ALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSTVTSWPSIVATGCTTTAT 504
 DB 421 ALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSTVTSWPSIVATGCTTTAT 480
 QY 505 PTGSGSVTSTSKTTATASKTSTTRS 530
 DB 481 PTGSGSVTSTSKTTATASKTSTTRS 506

RESULT 21

AAW55978
 ID AAW55978 standard; protein; 621 AA.

AC AAW55978;

XX 27-JUL-1998 (first entry)

DE Aspergillus awamori glucoamylase mutant 311-314 loop mutation.

XX Aspergillus awamori; glucoamylase; Aspergillus sp; mutant; fungal; food;

KW fructose; corn; sweetener; 1,4-alpha-D-glucan glucosyltransferase;

XX genetic engineering.

OS Synthetic.

XX Aspergillus awamori.

XX Key Location/Qualifiers

PT Misc-difference 311..319
 FT /note: "311-314 loop mutation"
 XX
 XX MO9803639-A1.
 XX
 XX 29-JAN-1998.
 XX
 XX 24-JUL-1997; 97WO-US012983.
 XX
 XX 24-JUL-1996; 96US-0022578P.
 XX
 XX 02-AUG-1996; 96US-0023077P.
 XX
 XX (IOWA) UNIV IOWA STATE RES FOUND INC.

XX Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzakto R;
 PI Ford C;
 XX
 XX MPI; 1998-120764/11.

PT Genetically engineered fungal glucoamylase - useful in, e.g. food
 PT industry for production of high fructose corn sweeteners.

PS Claim 7; Page; 97pp; English.

CC The present sequence represents a specifically claimed mutant
 CC glucoamylase from Aspergillus awamori (1,4-alpha-D-glucan
 CC glucosyltransferase). The present invention describes fungal glucoamylases
 CC (FG) comprising: a mutation pair Asn20Cys coupled with Ala27Cys forming a
 CC disulphide bond between the 2 members of the pair; and a 311-314loop or
 CC Ser41Ala mutation. FG can be used in industry for the production of high
 CC fructose corn sweeteners, while the glucose produced by glucoamylase can
 CC be crystallised or used in fermentation to produce organic products, e.g.
 CC citric acid, ascorbic acid, lysine, glutamic acid or ethanol for
 CC beverages and fuel. The mutations provide increased thermal stability,
 CC reduced isomaltose formation and increased pH optimum. N.B. The present
 CC sequence is not given in the specification but is derived from SEQ ID
 CC NO:1 as stated in the claim

XX Sequence 621 AA:

Query Match 93.8%; Score 2600.5; DB 2; Length 621;
 Best Local Similarity 98.2%; Pred. No. 1.1e-200;
 Matches 502; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 25 ATLDLWLSNEATVARTALINIGADGAWWSGADSGIVVSPSTNDPDYFTWTRDSGLVL 84
 DB 1 ATLDLWLSNEATVARTALINIGADGAWWSGADSGIVVSPSTNDPDYFTWTRDSGLVL 60
 QY 85 KTLVDLFRNGDTSLSLTIENTYISAQAIYVQGISNPSGDLSSGAGLGEKFNVDETAYTGSW 144
 DB 61 KTLVDLFRNGDTSLSLTIENTYISAQAIYVQGISNPSGDLSSGAGLGEKFNVDETAYTGSW 120
 QY 145 GRPORDPALRATAMIGFGQWLLDNGYTSATDIWPLVRNDLSYVAQYWNQGYDLMEW 204
 DB 121 GRPORDPALRATAMIGFGQWLLDNGYTSATDIWPLVRNDLSYVAQYWNQGYDLMEW 180
 QY 205 VNGSSFFTTIAVQHRALVEGSAFATAVAGSSCSWCDQAPILCYLQSFMTGSPFLANFDS 264
 DB 181 VNGSSFFTTIAVQHRALVEGSAFATAVAGSSCSWCDQAPILCYLQSFMTGSPFLANFDS 240
 QY 265 RSGKDNATLLGSIHTPDPEACDSTFQPCSPALANHEKVDSPRSITTLNDGLSDSEA 324
 DB 241 RSGKDNATLLGSIHTPDPEACDSTFQPCSPALANHEKVDSPRSITTLNDGLSDSEA 300
 QY 325 VAVGRYPEDTY-----YNGNPMFLCTLAABQLYDALYQMDKQSLVETDVSLEDFKALY 379
 DB 301 VAVGRYPEDTYVNGNPMFLCTLAABQLYDALYQMDKQSLVETDVSLEDFKALY 360
 QY 380 SDAATGTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNGSMSEQYDKSGEQLSARDL 439
 DB 361 SDAATGTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNGSMSEQYDKSGEQLSARDL 420
 QY 440 TWSYALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSSTVTSWPSIVATGCT 499

```
Db 421 TWSYAAALTNNRNNSVVPASWGETSASVPGTCAATSALIGTVSVTVTSPSIVATGCT 480
Qy 500 TTTATPTGSGSVTSTSKTTATATASKTSTTTRS 530
Db 481 TTTATPTGSGSVTSTSKTTATATASKTSTTSS 511

RESULT 22
AAB15183
ID AAB15183 standard; protein; 621 AA.
XX
AC AAB15183;
XX
DT 11-DEC-2000 (first entry)
XX
DE A. awamori mutant glucoamylase S30P/G137A/311-314 loop substitution.
XX
KM Glucoamylase; enzyme; carbohydrase; glucose;
XX 1,4-alpha-D-glucan glucosylhydrolase; mutcin; mutation.
XX
OS Aspergillus awamori.
XX Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 30 /note= "wild-type Ser substituted by Pro"
FT Misc-difference 137 /note= "wild-type Gly substituted by Ala"
FT Misc-difference 311..314 /note= "wild-type Tyr-Tyr-Asn-Gly substituted by Tyr-Asn-Gly-Asn-Gly-Asn-Ser-Gln-Gly"
FT FT
XX
XX WO200043504-A1.
XX
XX 27-JUL-2000.
XX
XX 10-JAN-2000; 2000WO-US000532.
XX
XX 22-JAN-1999; 99US-00236063.
XX
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
XX Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzalko R,
XX Ford C;
XX
XX WPI; 2000-514725/46.
XX
XX Fungal glucoamylase for selective production of glucose rather than alpha
XX -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
XX with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
XX Claim 42; Page; 160pp; English.
XX
XX Glucoamylase (1,4-alpha-D-glucan glucosylhydrolase; E.C. 3.2.1.3) is a
XX carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of
XX maltooligosaccharides, attacking alpha-(1,4)-, and alpha-(1,6)-glucosidic
XX bonds. The present invention relates to mutant glucoamylases, which have
XX increased thermostability, increased pH optimum and reduced isomaltose
XX formation. The mutant proteins are useful for the selective production of
XX glucose rather than alpha-1,6 linked disaccharide isomaltose. The present
XX sequence is one such mutant enzyme. Note: The present sequence is not
XX shown in the specification but is derived from the Aspergillus awamori
XX wild-type glucoamylase sequence given in pages 152-153 of the sequence
XX listing (SEQ ID 1)
XX
XX Sequence 621 AA;
XX
XX Query Match 93.5%; Score 2589.5; DB 3; Length 621;
XX Best Local Similarity 97.8%; Pred. No. 8.5e-200;
XX Matches 500; Conservative 1; Mismatches 5; Indels 5; Gaps 1;
XX
XX 25 ATLDWSLSEATVARTAILNNIGADGAWVSGADSGIVASBSTNDPDTFTWTRDSGLVL 84
```

```
Db 1 ATLDWSLSEATVARTAILNNIGADGAWVPGADSGIVVASPTSDNDPDTFTWTRDSGLVL 60
Qy 85 KTLVDFPRNGDTSILSTIENYISAQAIVOG1SNPSGDLSSGAGIGPKFENVDEATATG5W 144
Db 61 KTLVDFPRNGDTSILSTIENYISAQAIVOG1SNPSGDLSSGAGIGPKFENVDEATATG5W 120
Qy 145 GRPQDGPALRATAMTGFQGMLLDNGYSTAIDIWPLVRNDLSYVAQVWNGTGYDLME 204
Db 121 GRPQDGPALRATAMTAFQGMLLDNGYSTAIDIWPLVRNDLSYVAQVWNGTGYDLME 180
Qy 205 VNGSFFETIAVQHRALVEGSAFATAVGSSGWCDSQAPETLCYQSFWTGSFTLANPDS 264
Db 181 VNGSFFETIAVQHRALVEGSAFATAVGSSGWCDSQAPETLCYQSFWTGSFTLANPDS 240
Qy 265 RSGKDANTLLGSIHTEPDPEACDSTFQPCSPALANHKEVDSFFSIYTLNDGLSDSEA 324
Db 241 RSGKDANTLLGSIHTEPDPEACDSTFQPCSPALANHKEVDSFFSIYTLNDGLSDSEA 300
Qy 325 VAVGRYPEDTY----YNGNPWFLCTLAABEQLYDALYQMDKQSLVETDVSIDFFKALY 379
Db 301 VAVGRYPEDTYNNGNGNSQGNPWFCLTLAAAEQLYDALYQMDKQSLVETDVSIDFFKALY 360
Qy 380 SDAATGTYSSSSSTYSIVDAVKTFPADGFVSIYETHAASGWSBEQYDKSDGQLSARDL 439
Db 361 SDAATGTYSSSSSTYSIVDAVKTFPADGFVSIYETHAASGWSBEQYDKSDGQLSARDL 420
Qy 440 TWSYAAALTNNRNNSVVPASWGETSASVPGTCAATSALIGTVSVTVTSPSIVATGCT 499
Db 421 TWSYAAALTNNRNNSVVPASWGETSASVPGTCAATSALIGTVSVTVTSPSIVATGCT 480
Qy 500 TTTATPTGSGSVTSTSKTTATATASKTSTTTRS 530
Db 481 TTTATPTGSGSVTSTSKTTATATASKTSTTSS 511

RESULT 23
AAB15182
ID AAB15182 standard; protein; 621 AA.
XX
XX AAB15182;
XX
XX 11-DEC-2000 (first entry)
XX
XX A. awamori mutant glucoamylase N20C/A27C/S411A/311-314 loop substitution.
XX
XX Glucoamylase; enzyme; carbohydrase; glucose;
XX 1,4-alpha-D-glucan glucosylhydrolase; mutcin; mutation.
XX
XX Aspergillus awamori.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Disulfide-bond 20..27
XX Misc-difference 20 /note= "wild-type Asn substituted by Cys"
XX Misc-difference 27 /note= "wild-type Ala substituted by Cys"
XX Misc-difference 311..314 /note= "wild-type Tyr-Tyr-Asn-Gly substituted by Tyr-Asn-Gly-Asn-Gly-Asn-Ser-Gln-Gly"
XX Misc-difference 411 /note= "wild-type Ser substituted by Ala"
XX
XX WO200043504-A1.
XX
XX 27-JUL-2000.
XX
XX 10-JAN-2000; 2000WO-US000532.
XX
XX 22-JAN-1999; 99US-00236063.
XX
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.
```


XX Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzacko R;
PI Ford C;
XX
XX WPI: 2000-514725/46.
XX
XX Fungal glucanase for selective production of glucose rather than alpha
PT -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
PR with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
XX Claim 16; Page; 160pp; English.
XX
XX Glucoamylase (1,4-alpha-D-glucan glucohydrolase; E.C. 3.2.1.3) is a
CC carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of
CC maltooligosaccharides, attacking alpha-(1,4)- and alpha-(1,6)-glucosidic
CC bonds. The present invention relates to mutant glucanases, which have
CC increased thermostability, increased pH optimum and reduced isomaltose
CC formation. The mutant proteins are useful for the selective production of
CC glucose rather than alpha-1,6 linked disaccharide isomaltose. The present
CC sequence is one such mutant enzyme. Note: The present sequence is not
CC shown in the specification but is derived from the Aspergillus awamori
CC wild-type glucanase sequence given in pages 152-153 of the sequence
CC listing (SEQ ID 1)
XX
XX Sequence 621 AA;
SQ

Query Match 93.3%; Score 2584.5; DB 3; Length 621;
Best Local Similarity 97.7%; Pred. No. 2.1e-199;
Matches 499; Conservative 2; Mismatches 5; Indels 5; Gaps 1;

QY 25 ATLDLSWLSNEATVARTAILNINIGADGAWVSGADGIVVSPSTNDPDEFTYTRDSGLVY 84
DB 1 ATLDLSWLSNEATVARTAILNINIGADGAWVSGADGIVVSPSTNDPDEFTYTRDSGLVY 60
QY 85 KTLVDFRNGDTSLSLTENYISAQAIIVOGISNPSGDLSSGAGLGEPEKENVDEATVATGSM 144
DB 61 KTLVDFRNGDTSLSLTENYISAQAIIVOGISNPSGDLSSGAGLGEPEKENVDEATVATGSM 120
QY 145 GRPDRGPRALRATAMIGFGOMLDNGYTSATDITVPLVANDLSYVAQVWNOGYLMEB 204
DB 121 GRPDRGPRALRATAMIGFGOMLDNGYTSATDITVPLVANDLSYVAQVWNOGYLMEB 180
QY 205 VNGSSFFITAVOHRALVEGSAFATAVGSSCSWCDQAPELLCYLQSFMTGSFTLANFDS 264
DB 181 VNGSSFFITAVOHRALVEGSAFATAVGSSCSWCDQAPELLCYLQSFMTGSFTLANFDS 240
QY 265 RSGKDANTLLGSIHTFDPEACDSTFQPCSPALANHKEVDSFRSIYTLNDGLSDSEA 324
DB 241 RSGKDANTLLGSIHTFDPEACDSTFQPCSPALANHKEVDSFRSIYTLNDGLSDSEA 300
QY 325 VAVGRPEEDTY-----YNGNPWFLCTLAABEQYDALYQWDKGSLEVTDVSLDFPKALY 379
DB 301 VAVGRPEEDTYNGNGNSQGNPMWFLCTLAABEQYDALYQWDKGSLEVTDVSLDFPKALY 360
QY 380 SDAATGTYSSSSSTYSIYDAVTFADGFVSIYETHAASNGSMEQDYKSDGQLSARDI 439
DB 361 SDAATGTYSSSSSTYSIYDAVTFADGFVSIYETHAASNGSMEQDYKSDGQLSARDI 420
QY 440 TWSYALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSYVTSWPSIVATGCT 499
DB 421 TWSYALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSYVTSWPSIVATGCT 480
QY 500 TTTATPTGGSGVSTSKTATATKSTTTRS 530
DB 481 TTTATPTGGSGVSTSKTATATKSTTTRS 511

RESULT 24
ABP96630
ID ABP96630 standard; protein; 1095 AA.
XX AC ABP96630;
XX

DT 02-JUN-2003 (first entry)
XX
XX Alpha-amylase/glucanase fusion protein sequence SEQ ID NO:45.
DE
XX
XX Self-processing plant; plant; processing enzyme; alpha-amylase; grain;
XX pullulanase; alpha-glucosidase; glucose isomerase; glucanase;
XX mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;
XX maltodextrin; ethanol; fermentation; beverage; enzyme.
XX
XX Aspergillus shirousami.
OS
OS Synthetic.
XX
XX WO2003018766-A2.
XX
XX 06-MAR-2003.
XX
XX 27-AUG-2002; 2002WO-US027129.
XX
XX 27-AUG-2001; 2001US-0315281P.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Lanahan MB, Basu SS, Batie CJ, Chen W, Craig J, Kinkema M;
PI
XX
XX WPI: 2003-268420/26.
XX
XX N-PSDB; ACC44572.
XX
XX Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.
PT alpha-amylase, useful for producing plant to produce food products having
PT improved taste or fermentable substrates for ethanol.
XX
XX Claim 1; Page 107; 158pp; English.
XX

CC The present invention describes polynucleotides which encode processing
CC enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose
CC isomerase, or glucanase) that are optimised for expression in plants.
CC The polynucleotides encode mesophilic, thermophilic or hyperthermophilic
CC processing enzymes, which are activated under suitable conditions to act
CC upon the desired substrate. Also described are self-processing transgenic
CC plants and plant parts, e.g. grain, which express one or more of these
CC enzymes and have an altered composition that facilitates plant and grain
CC processing. Also described is a method (M) for converting starch to
CC starch-derived products in a transformed plant part (Tpp), by activating
CC the starch processing enzyme contained in it. Transgenic grain is useful
CC for preparing maltodextrin. A transformed plant (Tp) can be used to
CC produce food products having improved taste and to produce fermentable
CC substrates for ethanol and fermented beverages. (M) eliminates the need
CC to mill or physically disrupt the integrity of plant parts prior to
CC recovery of starch-derived products. The present sequence represents
CC alpha-amylase/glucanase fusion protein, which is given in the
CC exemplification of the present invention
XX
XX Sequence 1095 AA;
SQ

Query Match 91.1%; Score 2523; DB 6; Length 1095;
Best Local Similarity 92.9%; Pred. No. 4.7e-194;
Matches 461; Conservative 16; Mismatches 15; Indels 6; Gaps 2;

QY 13 VCTGLANVISKATIDSWLSNEATVARTAILNINIGADGAWVSGADGIVVSPSTNDPXY 72
DB 474 ICSS-----SKPATIDSWLSNEATVARTAILNINIGADGAWVSGADGIVVSPSTNDPXY 528
QY 73 FTTWTRDSGLVYKTLVDLFRNGDTSLSLTENYISAQAIIVOGISNPSGDLSSGAGLGEPEK 132
DB 529 FTTWTRDSGLVYKTLVDLFRNGDTSLSLTENYISAQAIIVOGISNPSGDLSSGAGLGEPEK 587
QY 133 FNVDETAATGSGWRGPRDGPALRATAMIGFGOMLDNGYTSATDITVPLVANDLSYVAQ 192
DB 588 FNVDETAATGSGWRGPRDGPALRATAMIGFGOMLDNGYTSATDITVPLVANDLSYVAQ 647
QY 193 YWNOGYDIMEEVENGSSFFITAVOHRALVEGSAFATAVGSSCSWCDQAPELLCYLQSF 252
DB 648 YWNOGYDIMEEVENGSSFFITAVOHRALVEGSAFATAVGSSCSWCDQAPELLCYLQSF 707

DR WPI; 2003-332729/31.
 DR N-PSDB; ABQ80353, ABQ80354.
 XX
 PT Novel isolated *Aspergillus fumigatus* polypeptide, useful in various
 PT industries such as those involved in the making of food and feed,
 PT detergents, textiles and detergents.
 XX
 PS Claim 17, Page 152-53; 169pp; English.
 XX
 The sequences given in ABB80164-87 show enzymatic proteins derived from
 CC *A. fumigatus*. These proteins display the catalytic activity of an enzyme
 CC such as tannase, cellulase, glucose oxidase, glucoamylase, phytase, beta-
 CC galactosidase, invertase, lipase, alpha-amylase, lactase,
 CC polygalacturonase or xylanase. Compositions comprising the tannase are
 CC useful for modulating the amount of compounds that comprise a gallate
 CC ester linkage in a composition. Compositions comprising cellulase are
 CC useful for modulating the amount of cellulose in a composition.
 CC Compositions comprising glucose oxidase are useful for modulating the
 CC amount of glucose or oxygen in a composition. Compositions comprising
 CC phytase are useful for modulating the amount of myo-inositol phosphates
 CC in a composition. Compositions comprising beta-galactosidase are useful
 CC for modulating the amount of lactose in a composition. Compositions
 CC comprising sucrase or invertase are useful for modulating the amount of
 CC sucrose in a composition. Compositions comprising lipase are useful for
 CC modulating the amount of glyceride in a composition. Compositions
 CC comprising alpha-amylases are useful for modulating the amount of
 CC starches or maltodextrins in a composition. Compositions comprising
 CC lactase are useful for modulating the amount of oxidized phenolic
 CC compounds in a composition. Compositions comprising polygalacturonases
 CC are useful for modulating the amount of high or low molecular weight
 CC polygalacturonic acid chains in a composition. Compositions comprising
 CC xylanases are useful for modulating the amount of xylan or xylo-oligomers
 CC in a composition. The *A. fumigatus* proteins and corresponding DNA's are
 CC useful in various industries such as those involved in the making of food
 CC and feed, beverages, textiles and detergents. The DNA's are useful to
 CC express recombinant enzymes for characterization, modification or
 CC industrial uses, to compare with the nucleotide sequence of *A. fumigatus*
 CC to identify duplicated genes of paralogs having the same or similar
 CC biochemical activity and/or function, to compare with nucleic acid
 CC sequence of other related or distant fungal organisms to identify
 CC potential orthologous enzyme genes, for selecting and making oligomers
 CC for attachment to a nucleic acid array for examination of expression
 CC patterns, and to raise anti-protein antibodies. The polypeptide having
 CC tannase activity increases the yield of tea liquor from tea leaves,
 CC improves the colour, flavour and health benefits of tea products,
 CC particularly an instant tea product. The polypeptide having cellulase
 CC activity enhances cleaning ability of detergent compositions
 CC
 XX
 SQ Sequence 631 AA;
 Query Match 66.8%; Score 1850.5; DB 6; Length 631;
 Best Local Similarity 66.3%; Pred. No. 3,6e-140;
 Matches 345; Conservative 65; Mismatches 97; Indels 13; Gaps 5;
 QY 21 ISKRAT--IDSWLSNEATVARTAIINNIGAGAMVSGAGSVIVASPTNDNPFYTWTR 78
 DB 23 LGRBARIGSLDSWIGTETVALNIGLIDNGAYASAKGIIITABSPISBPYYTWTTR 82
 QY 79 DSGLVAKTIVLDFRNGDTLSLTITENYISAQAIWVGISNBSGDLSSGAGLGEPEKENVDET 138
 DB 83 DALALVTKVIVLDFRNGNLGLQKIVITEVNSQAVLQTVSNBSGDLASG-GLAEPRKYNVDMT 141
 QY 133 AYNGSMGRPORDPARATAMIGFGOMLDNGTSTRATDVPWLVYVNDLSYVAQYNNQTC 198
 DB 142 AFTGAAGRPRDPARATLIDFGWMLIDNGYSYAVNNIWIYVNDLSYVSQYWSQSC 201
 QY 199 YDLMEVNGSSFTIAVQHRALVEGSAFATAVGSSCMCSQAPELICYQSFWTSFLL 258
 DB 202 FDMEEVNSMSTFTVAVQHRALVEGSAFATAVGSSCMCSQAPOLICTYQSFWTSYIN 261
 QY 259 ANFDSSRSGKADNTLLGSITFDPEAACDDSTFQPCSPRALANHKEVVDSPRSIYTLNDG 318
 DB 262 ANTGGRRSGKADNTLLASITFDPEAGCDDTTFQPCSPRALANHKEVVDSPRSVAINSG 321

QY 319 LSDSEAVAVGRYPEDTYNGNPMFLCTTLAAEQLYDALYQMDKQSLSEYTDVSLDFEKKAL 378
 DB 322 IFQGAANVASGRYPEDTYNGNPMFLCTTLAAEQLYDALYQMDKQSLSEYTDVSLDFEKKAL 381
 QY 379 YSDAATGYSSSSSTYSIVDAVKTFADGFSVIVETHAASNGSMSQYDKSDGEQLSARD 438
 DB 382 YSSAAVGTVAASSTSTFTDILNNAVKTYADGVYSIVQHANNNGSLSEQPKSKGLSLSARD 441
 QY 439 LMSYAAALLTANNRRSVVPASMGERSASVQCTCAATGYSVYTSWPSIYATG- 497
 DB 442 LMSYAAFLTANRRRGVVPAPMGASANSVSSCSMGSGATGYSTATATSWPSTLTSLS 501
 QY 498 -GTTT--TATPTGSGVSTSTKTAFA-----SKTSTT 528
 DB 502 PGSTTVGTSTTSGTATETACATPAVAVFNENIATTT 541
 RESULT 27
 AAB48170
 XX AAB48170 standard; protein; 624 AA.
 AC
 XX AAB48170;
 DT 02-APR-2001 (first entry)
 XX
 DE Thermoascus crustaceus glucoamylase polypeptide.
 XX
 XX Thermoascus crustaceus; glucoamylase; starch conversion; ethanol;
 KW maltose syrup; beverage; citric acid; ascorbic acid; detergent;
 KW thermostability; glucose.
 XX
 OS Thermoascus crustaceus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 80
 FT /note= "indicated incorrectly as Asp in the Fig 1, but
 FT correctly as Tyr in the sequence listing"
 FT
 XX
 PN MO200075296-A1.
 XX
 XX 14-DEC-2000.
 PD
 XX 02-JUN-2000; 2000MO-DK000301.
 PF
 XX 02-JUN-1999; 99DK-00000779.
 PR
 XX (NOVO) NOVO NORDISK AS.
 PA
 PI Nielsen BR, Kauppinen MS, Nielsen RI;
 XX
 DR WPI; 2001-071066/08.
 DR N-PSDB; AAC84443.
 XX
 PT Novel Thermoascus crustaceus glucoamylase polypeptide useful in a
 PT continuous starch conversion process, detergents and for producing
 PT oligosaccharides, speciality syrups, ethanol for fuel or drinking, and
 PT beverages.
 PT
 XX
 PS Claim 4; Fig 1; 61pp; English.
 XX
 This representsa Thermoascus crustaceus glucoamylase polypeptide. The
 CC glucoamylase polypeptide can be expressed by standard recombinant
 CC methodology and is useful in a continuous starch conversion process, for
 CC producing oligosaccharides, speciality syrups such as maltose compounds,
 CC ethanol for fuel or drinking ethanol, beverages, and organic compounds
 CC such as citric acid, ascorbic acid, lysine or glutamic acid. It is also
 CC useful in detergents such as laundry detergent compositions, dish wash
 CC compositions and/or hard surface cleaning compositions. The T. crustaceus
 CC glucoamylase has higher thermostability than *Aspergillus niger* GI
 CC glucose reversion tendency

KW dextrose syrup; syrup; ethanol; fuel; beverage; citric acid;
 KM ascorbic acid; lysine; glutamic acid.
 XX Talaromyces emersonii.
 OS WO9928448-A1.
 XX
 XX
 PD 10-JUN-1999.
 XX
 PF 26-NOV-1998; 98MO-DK000520.
 XX
 XX 26-NOV-1997; 97US-00979673.
 PR 30-DEC-1997; 97DK-00001557.
 PR 30-JUN-1998; 98US-00107657.
 PR 10-JUL-1998; 98DK-00000925.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Nielsen BR, Nielsen RI, Lehmbeck J;
 XX
 DR WPI; 1999-404822/34.
 XX
 XX New glucamylase obtained from Talaromyces emersonii.
 PS Claim 8; Page 57-59; 79pp; English.
 XX
 XX The present sequence represents a glucamylase enzyme obtained from
 CC Talaromyces emersonii. The glucamylase enzymes have high thermal
 CC stability so that a saccharification process may be carried out within a
 CC shorter period of time or the process may be carried out using a lower
 CC enzyme dosage. The glucamylase enzymes can be used for saccharifying
 CC starch hydrolyase for converting starch or partially hydrolysed starch
 CC into a syrup containing dextrose. They can be used for producing
 CC oligosaccharides, specialty syrups, ethanol for fuel, beverages or
 CC organic compounds such as citric acid, ascorbic acid, lysine or glutamic
 CC acid
 XX
 SQ Sequence 591 AA;
 Query Match 60.4%; Score 1674.5; DB 2; Length 591;
 Best Local Similarity 62.3%; Pred. No. 5.2e-126;
 Matches 314; Conservative 73; Mismatches 102; Indels 15; Gaps 3;
 QY 26 TLDSWLSNEATVARTALINNIGADGAWVSGADSGIVVSPSTDPYFYTWRTDGLVTK 85
 DB 4 SLDSFLATETPIALQGVINNIGPNAGADVAGASGIVVASSRSDPFTFSWTRDALYTK 63
 QY 86 TLVDLFRNGDTSLLSTIENYISAQAIVOGISNPSGDLSSGAGGEPKFVNDETAYTGSNG 145
 DB 64 YLVDAENRGNKOLEQITQOYISAQAVQRTISNPSGLSTG-GIGEPKFVNNETAFITGPMG 122
 QY 146 RPRQDGPALRATMIGFGOMLNDNGTSTRTDIWPLVNRNDSLYVAQYNNQTCYDIWEEY 205
 DB 123 RPRQDGPALRATMIGFGOMLNDNGTSTRTDIWPLVNRNDSLYVAQYNNQTCYDIWEEY 182
 QY 206 NSSSFTIIVORHALVEGSAFATAVSSGSCWCSOAPETLCYOSQWTSFTLIANF-DSS 264
 DB 183 EGSSFTIIVORHALVEGSAFATAVSSGSCWCSOAPETLCYOSQWTSFTLIANF-DSS 242
 QY 265 RSGDKANTLLGSIHTDPEAACDSTFOPCSPRALNNHKEVDSFRTIYLANDGLDSEK 324
 DB 243 RSGDKVNSILGSIHTDPAAGCDDSTFOPCSARALNNHKEVDSFRTIYLANDGLDSEK 302
 QY 325 VAVGRPEPTTYNGNFWFLCTLAARQYDALYQMDKQSLVYTDVSLDFKALYSDAAT 384
 DB 303 VAVGRPEPTTYNGNFWFLCTLAARQYDALYQMDKQSLVYTDVSLDFKALYSDAAT 362
 QY 385 GTVSSSSSTYSIVDVVKTFPADGFVSIETHAASNGSMSEQYDKSGEOLSAADLTWSTYA 444
 DB 363 GTVSSSTTFNDIISAVQITGDBGISLIVERTYSDSGLTEQFSRTDGTPLSASALTWSTYA 422
 QY 445 ALLTANNRRNSVVPASWGETSASSVPGTCATSAIGTYSVTYTSWPSIVATGTTTAT 504

DB 423 SLTTSARRQSVVPASWGESSASSVLAVCSATSGATGPTATNTWPS----- 470
 QY 505 PTGSGSVTSTSKTATYASTKSTTT 528
 DB 471 -SGSGSTTSSAPCTPTTSVAVT 493
 RESULT 30
 AAM51596
 ID AAM51596 standard; protein; 630 AA.
 XX
 AC AAM51596;
 XX
 DT 07-FEB-2002 (first entry)
 XX
 DE Thielavia terrestris ATCC 20627 glucamylase.
 XX
 XX Glucamylase; fungus; thermophilic; dextrose; fructose syrup; beer;
 KW alcohol.
 XX
 OS Thielavia terrestris.
 XX
 PN US6309872-B1.
 PD 30-OCT-2001.
 XX
 PF 01-NOV-2000; 2000US-00704449.
 XX
 PR 01-NOV-2000; 2000US-00704449.
 XX
 PA (NOVO) NOVOZYMES BIOTECH INC.
 XX
 PI Rey MW, Golightly EJ;
 XX
 DR WPI; 2002-033282/04.
 DR N-PSDB; ABA01139.
 XX
 PT New Thielavia terrestris polypeptide with glucamylase activity, useful
 PT in the production of dextrose and fructose syrups, beer with low
 PT carbohydrate content and alcohol from fermentation of raw starch.
 PS
 PS Claim 5; Fig 1; 42pp; English.
 XX
 XX The invention relates to an isolated Thielavia terrestris polypeptide
 CC with glucamylase activity. Thielavia terrestris is a thermophilic
 CC filamentous fungus that can grow at low pH of 4.5 and elevated
 CC temperatures of 40-45 degreesC. The glucamylase polypeptide catalyses
 CC endo-hydrolysis of 1,6-alpha-D-glucoside linkages at points of branching
 CC chains in 1,4-linked alpha-D-glucose residues. The polypeptide may be
 CC used in the production of dextrose and fructose syrups, beer with low
 CC carbohydrate content and alcohol from fermentation of raw starch. The
 CC present sequence is the polypeptide of the invention
 CC
 SQ Sequence 630 AA;
 Query Match 58.3%; Score 1615.5; DB 5; Length 630;
 Best Local Similarity 58.9%; Pred. No. 3.3e-121;
 Matches 315; Conservative 83; Mismatches 112; Indels 25; Gaps 7;
 QY 6 LIALSLGVTGLANVISKRA-----TLDSWLSNEATVARTALINNIGADGAWVSGA 56
 DB 6 LIGLALLALPALAGHPASRVREGEVYKRSVDSFATSPALSNLNCIGSGCHASGV 65
 QY 57 DSGIIVASPTNDPDPFYTWRTDGLVTLVDLPNG-DTSLSTIENYISAQAIVOGI 115
 DB 66 ASGIIVASPTNDPDPFYTWRTDGLVTLVDLPNG-DTSLSTIENYISAQAIVOGI 125
 QY 116 SNPSGDLSSGAGGEPKFVNDETAYTGSNGRPRQDGPALRATMIGFGOMLNDNGTSTA 175
 DB 126 SNPSGDLSSGAGGEPKFVNDQFTGANGRPRQDGPALRATMIGFGOMLNDNGTSTA 185
 QY 176 TDIWPLVNRNDSLYVAQYNNQTCYDIWEEYSSFTIIVORHALVEGSAFATAVSSGS 235

Db 186 SSIWPIKNDLAVAO--NNTGFDLMEVSGSSFTVNAOHRLVGCALALATSLGSCS 243
QY 236 WCDSDAEILCYLQSFMTGS--FILANFDSSRGKDANTLLGSIHTPDPEACDSTFOP 293
Db 244 ACSAVAPQILCFLOQSFSPSSGYILAN----STAKDANTLLGSIHTPDPAAGCDATFOP 299
QY 294 CSRRALANHEVVDSPFSITLNDGLSDSEAVAVGRPEPTTYNGNWFELCTLAAROLY 353
Db 300 CSRRALANHEVVDSPFSITLNDGLSDSEAVAVGRPEPTTYNGNWFELCTLAAROLY 359
QY 354 DALYOMDQGLSEVTDVSLDFPKALYSDAATGTSSSSSTYSIVDAVKTFADGFSIVE 413
Db 360 DALYOMDQGLSEVTDVSLDFPKALYSDAATGTSSSSSTYSIVDAVKTFADGFSIVE 419
QY 414 THAASNGSMSEQYDKSDGQSLARDLTWSYAAALLTANNRNSVVPASWGETSASVPGTC 473
Db 420 QYAGTNGSLSEQFSKTNGEPLSAVDLTWSYAAFLTAARAGVYVPSWGAASANSVPAQC 479
QY 474 AATSATGTSVTVTSPSIATGCTTTATPTGSSGVSTSKTATASKTSTTT 528
Db 480 SATSVGSSYTSATATSPFP-----SQTPTSSTAGS-SPASSTTATATACSTPT 527

RESULT 31

AAR71034

ID AAR71034 standard; protein; 626 AA.

XX AAR71034;

XX 25-MAR-2003 (revised)

DT 18-MAR-1996 (first entry)

XX N. crassa glucanase.

XX Neurospora crassa; glucanase gene; gla-1; promoter;

KM expression construct; primer.

XX Neurospora crassa.

OS Key Location/Qualifiers

XX Peptide 1..19

FT Protein /note= "signal peptide"

FT /note= "mature gla-1"

FT Misc-difference 27..28

FT /note= "propeptide processing site"

FT /note= "propeptide processing site"

XX W09505474-A2.

XX 23-FEB-1995.

XX 15-AUG-1994; 94WO-GB001789.

XX 13-AUG-1993; 93GB-00016883.

XX (UYLE-) UNIV LEBDS.

XX Radford A, Parish JH;

XX WPI; 1995-098771/13.

XX N-PSDB; AA084689, AA084695.

XX Regulated glucanase promoter - useful for producing heterologous

XX polypeptide(s) in filamentous fungi.

XX Claim 1, Fig 1; 36pp; English.

XX This sequence represents N. crassa glucanase. The gene encoding this

XX sequence has high expression and is one of the major secreted proteins of

XX N. crassa when starch-induced. The ORF of this gene and the promoter

XX region can be attached in frame to a foreign gene thereby introducing

CC foreign proteins into the N. crassa secretory pathway. The 5' primer
CC (AA084690) encompassed the unique PvuII site at posn. 2163 of the
CC encoding sequence and the 3' primer (AA084691) cont. an MroI site
CC hybridizes at the 3' end of the gla gene. The 5' upstream PCR fragment
CC was amplified and cloned into the SmaI site in a pMB1 193 vector. The
CC clone was named pMro. The remainder of the gla gene was inserted by
CC digestion of the glucanase clone pGla-XhoI (AA084692). This plasmid
CC contains the entire gla gene however the downstream unsequenced and non-
CC transcribed area was deleted. pGla-XhoI was digested with SacI and PvuII
CC and the fragment ligated into the SacI/PvuII sites of pMro. The SacI
CC site of pGla-XhoI was derived from the linker and not from the coding
CC region of glucanase consequently, no glucanase sequence was deleted
CC (see AA084693, pGla-MroI). In an attempt to increase transcription the
CC efficiency, 1575 bp were deleted from the encoding sequence, creating the
CC plasmid pGE (plasmid glucanase, EcoRI). Deposits of plasmids pGla-Xho
CC (AA084692), pGla-MroI (AA084693), pGE (AA084694) have been made and the
CC deposition details are to be added to the patent application. (Updated on
CC 25-MAR-2003 to correct pN field.)
XX
XX Sequence 626 AA;

Query Match 56.9%; Score 1577; DB 2; Length 626;

Best Local Similarity 57.6%; Pred. No. 4,1e-118;

Matches 306; Conservative 85; Mismatches 130; Indels 10; Gaps 6;

QY 2 SFRSLALSGLVCTGLANVISKRAITDLSNENATVARTAILNNIGADGAMVSGADGIV 61

Db 13 AFQAVLGLPDLPEKHSIDIRK-SVDSYIQTETPIAKNKLNCNIGASGRASGASGVV 71

QY 62 VASPTNDPDYPTVTRDGLVKTVDLFPNG-DISLSTIENYISAQIVGGINPSG 120

Db 72 VASPSKSPDYMTWTRDALVTKLVDEFNDYNTLTIOAVAAQKLGVSNSPG 131

QY 121 DLSGAGLGEPEKENVDETAAYTGSWGRPORDGALRAFAMIGFQMLDNGYSTADIW 180

Db 132 SLNSGAGLGEPEKENVDETAAYTGSWGRPORDGALRAFAMIGFQMLDNGYSTADIW 191

QY 181 PLVRNDLSYVAQYNTGTYDLMEEVNGSFFTYAVOHRALVBSGAPATVAGSSCWDSO 240

Db 192 PLVRNDLSYVAQYNTGTYDLMEEVNGSFFTYAVOHRALVBSGAPATVAGSSCWDSO 251

QY 241 APRILCYLQSFMTGS--FILANFDSSRGKDANTLLGSIHTPDPEACDSTFOPCSPRAL 299

Db 252 APRILCYLQSFMTGS--FILANFDSSRGKDANTLLGSIHTPDPEACDSTFOPCSPRAL 311

QY 300 ANHEVVDSPFSIYTLNDGLSDSEAVAVGRPEPTTYNGNWFELCTLAAROLYDALYOW 359

Db 312 ANHEVVDSPFSIYTLNDGLSDSEAVAVGRPEPTTYNGNWFELCTLAAROLYDALYOW 370

QY 360 DKQSLLEVTDVSLDFPKALYSDAATGTSSSSSTYSIVDAVKTFADGFSIVETHAASN 419

Db 371 DKQSLLEVTDVSLDFPKALYSDAATGTSSSSSTYSIVDAVKTFADGFSIVETHAASN 430

QY 420 GSWSEQYDKSDGQSLARDLTWSYAAALLTANNRNSVVPASWGETSASVPGCATSAL 479

Db 431 GSWSEQYDKSDGQSLARDLTWSYAAALLTANNRNSVVPASWGETSASVPGCATSAL 490

QY 480 GTSVTVTSPSIATGCTTTATPTGSSGS-----VTSTSKTATASKT 524

Db 491 GTSVTVTSPSIATGCTTTATPTGSSGS-----VTSTSKTATASKT 541

RESULT 32

AAB18823

ID AAB18823 standard; protein; 581 AA.

XX AAB18823;

XX 08-FEB-2001 (first entry)

XX Amino acid sequence of a glucanase polypeptide.

XX Glucanase; protein production; promoter; hormone; receptor; antibody.

```

XX OS Fusarium venenatum.
XX FH Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /note= "signal peptide"
XX PN MO200056900-A2.
XX PD 28-SEP-2000.
XX PF 22-MAR-2000; 2000MO-US007815.
XX PR 22-MAR-1999; 99US-00274449.
XX PA (NOVO ) NOVO NORDISK BIOTECH INC.
XX PI Berka RM, Rey MW, Brown K, Brown SH;
XX DR WPI; 2000-638265/61.
XX DR N-PSDB; AAA75945.
XX PT Promoters useful for expressing heterologous genes and producing
XX PT polypeptides such as hormones, receptors, antibodies or enzymes in a
XX PT fungal cell.
XX PS Example 8; Fig 1A-F; 104pp; English.
XX CC The present sequence represents a glucoamylase polypeptide. The promoter
XX CC sequence of this gene is useful for producing a polypeptide, preferably a
XX CC hormone, receptor, antibody, a reporter or an enzyme selected from
XX CC oxidoreductase, transferase, hydrolase, lyase, isomerase or ligase, in
XX CC particular aminopeptidase, amylase, carboxypeptidase, carboxypeptidase,
XX CC catalase, cellulase, chitinase, cutinase, cyclodextrin glycosyl
XX CC transferase, deoxyribonuclease, esterase, alpha-galactosidase, beta-
XX CC galactosidase, glucoamylase, alpha-glucosidase, beta-glucosidase,
XX CC invertase, laccase, lipase, mannosidase, mutanase, oxidase, a
XX CC pectinolytic enzyme, peroxidase, phytase, polyphenoloxidase, proteolytic
XX CC enzyme, ribonuclease, transglutaminase or xylanase, in fungal host cells
XX SQ Sequence 581 AA;

Query Match 51.4%; Score 1425.5; DB 3; Length 581;
Best Local Similarity 52.2%; Pred. No. 6.2e-106;
Matches 273; Conservative 81; Mismatches 146; Indels 23; Gaps 6;

QY 9 LSGLVCTGL--ANVI---SKRATLDLWLSNEATVARTAIINIGADGAWVSGADSGIIVA 63
DB 6 LYLVAASALMQGVASPSKDNLSERIDKQADISIKVLANIGADGKRAQGAAPAVVA 65
QY 64 SPSTNDPDIYFYTWTRDSGLVLTVDLFRNGDTSLSTINNYISAQAIIVGGINSPGDL 123
DB 66 SPKEDPDVYIYTWTRDSGLVLTVDLFRNGDTSLSTINNYISAQAIIVGGINSPGDL 125
QY 124 SGAGLEPKRNVDETYTSGMGRPORGPALRATATMGFGOMLLDNGYSTATDIWPLV 183
DB 126 SG-GLSEPKRNVDETYTSGMGRPORGPALRATATMGFGOMLLDNGYSTATDIWPLV 184
QY 184 RNDLSVYAQYQWNTGYDLMEEVNGSSFFITAVOHRALVESSAATAVAGSSCSMDSQAP 243
DB 185 EKDLATTTKFMNRTGVDLMEVNGSSFFITASGRALVESSAATAVAGSSCSMDSQAP 244
QY 244 ILCTYLOSFWTGSTILANFD--SSRSGKANTLIGSITHTPEPAACDDSTQPCSPRALAN 301
DB 245 VLFCLQTFMVGQYVDSINIVKDGKGLDVNSIISHTFPNSKCTDSTQPCSPRALAN 304
QY 302 HKEVVSFRSIIYTLNDGLDSEAVAVGRYPEDTYNGNPNFLCTLAABEOLYALYOMDK 361
DB 305 HKAVVDSFRSIIYTLNDGLDSEAVAVGRYPEDTYNGNPNFLCTLAABEOLYALYOMDK 364
QY 362 QGSLLEVTVSLDFPKALYSDAATGYSSSSSYSSIVDAVATPDGFSIVETHAASNGS 421
DB 365 LGAIVTVDDVSLSPFKKIDVPRKVSQGYAKKRTYKEIIKAAKTYADGVAVAVQYTPPDGS 424

```

```

QY 422 MSEQYDKSDGEBOLASDILTWYAAITLANNRRNSVVPASGERTSASSVGTCAATSAIGT 481
DB 425 LMEQFDKSGAPRPSAHLTWYVAFVATERRDGIIISPWGESSAKVAVQCAAPACDT 484
QY 482 -----YSSVTYTSWPSIVATGTTTAA-----TPTSGG 509
DB 485 TTFPSVKNVQVSSDQKVVYVGSVTLELNSMSPDDGIALTFPSSSG 527

RESULT 33
AAU79444
ID AAU79444 standard; protein; 581 AA.
XX AAU79444;
XX AC AAU79444;
XX DT 02-JUL-2002 (first entry)
XX DE Fusarium venenatum glucoamylase protein.
XX KM Glucoamylase; enzyme; promoter; foreign gene expression.
XX OS Fusarium venenatum.
XX PN US6361973-B1.
XX PD 26-MAR-2002.
XX PF 22-MAR-2000; 2000US-00534407.
XX PR 22-MAR-1999; 99US-00274449.
XX PR 22-JUL-1999; 99US-0145339P.
XX PA (NOVO ) NOVOZYMES BIOTECH INC.
XX PI Berka RM, Rey MW, Brown K, Brown SH;
XX DR WPI; 2002-350392/38.
XX DR N-PSDB; ANK49157.
XX PT Novel Fusarium venenatum promoters, useful for expressing genes in a
XX PT fungal cell.
XX PS Example 7; Fig 1; 57pp; English.
XX CC The invention relates to a novel promoter sequence from the F. venenatum
XX CC glucoamylase gene, a subsequence of the promoter that retains promoter
XX CC activity or a nucleic acid sequence that hybridizes under medium
XX CC stringency conditions with the promoter. Also included are a method of
XX CC producing a polypeptide comprising cultivating a fungal host cell in a
XX CC medium, where the fungal host cell comprises a first nucleic acid
XX CC sequence encoding the polypeptide operably linked to a second nucleic
XX CC acid sequence comprising a promoter foreign to the first nucleic acid
XX CC sequence, where the promoter comprises a nucleic acid sequence selected
XX CC from the F. venenatum glucoamylase promoter, a subsequence of the
XX CC promoter that retains promoter activity and a nucleic acid sequence that
XX CC hybridizes under medium stringency conditions with the promoter and
XX CC isolating the polypeptide from the cultivation medium. The promoter is
XX CC useful for expressing genes in a fungal cell. Expression of a lipase
XX CC reporter gene in Fusarium venenatum when operably linked to the promoter
XX CC showed higher levels of lipase activity than a Fusarium oxysporum trypsin
XX CC promoter. Also disclosed are two novel genes designated Daria (a secreted
XX CC protein) and Quim (a vacuolar associated protein). The present sequence
XX CC is the F. venenatum glucoamylase protein sequence
XX SQ Sequence 581 AA;

Query Match 51.4%; Score 1425.5; DB 5; Length 581;
Best Local Similarity 52.2%; Pred. No. 6.2e-106;
Matches 273; Conservative 81; Mismatches 146; Indels 23; Gaps 6;

QY 9 LSGLVCTGL--ANVI---SKRATLDLWLSNEATVARTAIINIGADGAWVSGADSGIIVA 63

```


KW glucose; oxygen; myo-inositol phosphate; lactose; tea leaf; sucrose;
 KW glyceride; starch; maltodextrin; oxidized phenolic compound;
 KW polygalacturonic acid chain; xylan; xylo-oligomer; food; feed; beverage;
 KW textile; tea liquor; cleaning ability.
 XX
 OS Aspergillus fumigatus.
 PN WO2003012071-A2.
 XX
 PD 13-FEB-2003.
 XX
 PF 05-AUG-2002; 2002WO-US024842.
 XX
 PR 03-AUG-2001; 2001US-0309870P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 PI Jiang B, Storms R, Roemer T, Buesey H,
 DR N-PSDB; ABO80331, ABO80332.
 XX
 PT Novel isolated Aspergillus fumigatus polypeptide, useful in various
 PT industries such as those involved in the making of food and feed,
 PT beverages, textiles and detergents.
 PS
 PS Claim 17; Page 100-01, 169pp; English.
 XX
 CC The sequences given in ABB80164-87 show enzymatic proteins derived from
 CC A. fumigatus. These proteins display the catalytic activity of an enzyme
 CC such as tannase, cellulase, glucose oxidase, glucoamylase, phytase, beta-
 CC galactosidase, invertase, lipase, alpha-amylase, lactase,
 CC polygalacturonase or xylanase. Compositions comprising the tannase are
 CC useful for modulating the amount of compounds that comprise a galate
 CC ester linkage in a composition. Compositions comprising cellulase are
 CC useful for modulating the amount of cellulose in a composition.
 CC Compositions comprising glucose oxidase are useful for modulating the
 CC amount of glucose or oxygen in a composition. Compositions comprising
 CC phytase are useful for modulating the amount of myo-inositol phosphates
 CC in a composition. Compositions comprising beta-galactosidase are useful
 CC for modulating the amount of lactose in a composition. Compositions
 CC comprising sucrase or invertase are useful for modulating the amount of
 CC sucrose in a composition. Compositions comprising lipase are useful for
 CC modulating the amount of glyceride in a composition. Compositions
 CC comprising alpha-amylase are useful for modulating the amount of
 CC starches or maltodextrins in a composition. Compositions comprising
 CC lactase are useful for modulating the amount of oxidized phenolic
 CC compounds in a composition. Compositions comprising polygalacturonases
 CC are useful for modulating the amount of high or low molecular weight
 CC polygalacturonic acid chains in a composition. Compositions comprising
 CC xylanases are useful for modulating the amount of xylan or xylo-oligomers
 CC in a composition. The A. fumigatus proteins and corresponding DNA's are
 CC useful in various industries such as those involved in the making of food
 CC and feed, beverages, textiles and detergents. The DNA's are useful to
 CC express recombinant enzymes for characterization, modification or
 CC industrial uses, to compare with the nucleotide sequence of A. fumigatus
 CC to identify duplicated genes of paralogs having the same or similar
 CC biochemical activity and/or function, to compare with nucleic acid
 CC sequence of other related or distant fungal organisms to identify
 CC potential orthologous enzyme genes, for selecting and making oligomers
 CC for attachment to a nucleic acid array for examination of expression
 CC patterns, and to raise anti-protein antibodies. The polypeptide having
 CC tannase activity increases the yield of tea liquor from tea leaves,
 CC improves the colour, flavour and health benefits of tea products,
 CC particularly an instant tea product. The polypeptide having cellulase
 CC activity enhances cleaning ability of detergent compositions
 CC
 XX
 SO Sequence 704 AA:

Query Match 50.8%; Score 1408.5; DB 6; Length 704;
 Best Local Similarity 52.0%; Pred. No. 1.9e-104;
 Matches 268; Conservative 85; Mismatches 139; Indels 23; Gaps 5;

QY 27 LDSWLSNEATVARTAIINNIGADGAWVSGADSGIVVASFSTNDPFIYTWTDGSLVLT 86
 DB 118 LVSWLAQETSVALDGVNINVGPNKAKATGASGIIIIASPSQNPDIYYTWTDAAITVYK 177
 QY 87 LYDLF-RNGDTSLSSTIENYISAQAVGCSINPSGSLSGALGEPKFNVDERATYVSGMG 145
 DB 178 LVGSFPAADHPAIQRIIEDYVESQAHLOTVSNPSGLSSG-GLGEKFLVVDGSAFPGSWG 236
 QY 146 RPPORDPALRATAMIGFGQWMLDNGYTSATATDIWPELVNNDLSYVQYNNQTSYDLMEEV 205
 DB 237 RPPSDGPARATTLISYALMIDNGYTSVTSIWIPIQNDLSYLFEPNNSFTDLMEEV 296
 QY 206 NGSSFFTTAVGQRAVVEGSAFATVAGSSCSWCDSQAPILLCYLOSFWTGSFIILANFDSR 265
 DB 297 RGSFFTTAVGQRAVVEGSAFATVAGSSCSWCDSQAPILLCYLOSFWTGSFIILANFDSR 356
 QY 266 SGKDANTLLGSIHTPPEAACDDSTFQPCSPRALANHKVNDGFRSITYLNDGSLSEAV 325
 DB 357 SGKDANSILGIHTPDPNAGCDGQTFQPCSDRALSHKEVDSFRSLYPMADIPOQNAV 416
 QY 326 AVGRYPEDTYNGNPMFLCTLLAAEQYDALYQMDKQSLSEYTVDELDFPKALYSPAATG 385
 DB 417 AVGRYPEDTYNGNPMFLCTLLAAEQYDALYQMDKQSLSEYTVDELDFPKALYSPAATG 476
 QY 386 TYSSSSSTYSIVDAVKTPEADGFVSIVETHAASNGSMSBQYKSDGQSLASDRLTWSYAA 445
 DB 477 TYAKDITFASISAAVROVADRFRLVQKYTPENGALAQFSRYDGSPLSAQDLTWSYTS 516
 QY 446 LFTA-----NRRNSVVPASWGETSASVPGTAATSAIGTYSVTVTSNP----- 491
 DB 537 FLTAQVARRHAIINPSASHIQPLISNATVYTAALPOVCTPPSARGPYQPVKRIKPRECLS 596
 QY 492 --SIVATGTTTTAAPTGS-----GSVTSKTTA 519
 DB 597 PSTVAVRNVIAATVIGEDIFLVGSIIPALGEMDA 631

RESULT 36
 ABB80174
 ID ABB80174 standard; protein; 620 AA.
 XX
 AC ABB80174;
 XX
 DT 11-AUG-2003 (first entry)
 XX
 DE A. fumigatus AFGLA2.
 XX
 XX Enzyme; tannase; cellulase; glucose oxidase; glucoamylase; phytase;
 KW beta-galactosidase; invertase; lipase; alpha-amylase; lactase;
 KW polygalacturonase; xylanase; galate ester linkage; detergent; cellulose;
 KW glucose; oxygen; myo-inositol phosphate; lactose; tea leaf; sucrose;
 KW glyceride; starch; maltodextrin; oxidized phenolic compound;
 KW polygalacturonic acid chain; xylan; xylo-oligomer; food; feed; beverage;
 KW textile; tea liquor; cleaning ability.
 XX
 OS Aspergillus fumigatus.
 PN WO2003012071-A2.
 XX
 PD 13-FEB-2003.
 XX
 PF 05-AUG-2002; 2002WO-US024842.
 XX
 PR 03-AUG-2001; 2001US-0309870P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 PI Jiang B, Storms R, Roemer T, Buesey H,
 DR N-PSDB; ABO80331, ABO80332.
 PT Novel isolated Aspergillus fumigatus polypeptide, useful in various

PT industries such as those involved in the making of food and feed,
PT beverages, textiles and detergents.

XX Claim 17, Page 123-24; 169pp; English.

CC The sequences given in ABB80164-87 show enzymatic proteins derived from
CC A. fumigatus. These proteins display the catalytic activity of an enzyme
CC such as tannase, cellulase, glucose oxidase, glucoamylase, phytase, beta-
CC galactosidase, invertase, lipase, alpha-amylase, lactase,
CC polygalactonase or xylanase. Compositions comprising the tannase are
CC useful for modulating the amount of compounds that comprise a gallate
CC ester linkage in a composition. Compositions comprising cellulase are
CC useful for modulating the amount of cellulose in a composition.
CC Compositions comprising glucose oxidase are useful for modulating the
CC amount of glucose or oxygen in a composition. Compositions comprising
CC phytase are useful for modulating the amount of myo-inositol phosphates
CC in a composition. Compositions comprising beta-galactosidase are useful
CC for modulating the amount of lactose in a composition. Compositions
CC comprising sucrose or invertase are useful for modulating the amount of
CC sucrose in a composition. Compositions comprising lipase are useful for
CC modulating the amount of glyceride in a composition. Compositions
CC comprising alpha-amylases are useful for modulating the amount of
CC starches or maltodextrins in a composition. Compositions comprising
CC lactase are useful for modulating the amount of oxidized phenolic
CC compounds in a composition. Compositions comprising polygalactonases
CC are useful for modulating the amount of high or low molecular weight
CC polygalacturonic acid chains in a composition. Compositions comprising
CC xylanases are useful for modulating the amount of xylan or xyl-o-oligomers
CC in a composition. The A. fumigatus proteins and corresponding DNA's are
CC useful in various industries such as those involved in the making of food
CC and feed, beverages, textiles and detergents. The DNA's are useful to
CC express recombinant enzymes for characterization, modification or
CC industrial uses, to compare with the nucleotide sequence of A. fumigatus
CC to identify duplicated genes of paralogs having the same or similar
CC biochemical activity and/or function, to compare with nucleic acid
CC sequence of other related or distant fungal organisms to identify
CC potential orthologous enzyme genes, for selecting and making oligomers
CC for attachment to a nucleic acid array for examination of expression
CC patterns, and to raise anti-protein antibodies. The polypeptide having
CC tannase activity increases the yield of tea liquor from tea leaves,
CC improves the colour, flavour and health benefits of tea products,
CC particularly an instant tea product. The polypeptide having cellulase
CC activity enhances cleaning ability of detergent compositions

CC Sequence 620 AA;

Query Match 49.8%; Score 1380.5; DB 6; Length 620;

Best Local Similarity 53.8%; Pred. No. 2.9e-102; Mismatches 267; Conservative 75; Indels 23; Gaps 3;

QY 13 VCTGLANVSKRAT-----LDSSLNEATVARTALINNIGADGAWVGADS 58
DB 6 IAVGIASVILGHLATPSSAKELAARSPAAANSRLVKEGIFATESLALANGINAPPTAA 65
QY 59 GIVVASPESTNDPFFYTWTRDSGLVLTVDLPFNQGT-----SLISTENTYSAQAI 111
DB 66 GLTIASPTIQNPDPFYTWTRDAALTFKGLVDIFIGCGTFFIVNLDGETHIQDIYISSQAV 125
QY 112 VQGISNSGDISAGGEGPEKENVDETAATYSGWGRPORDGPALRATMIGCGQLNGY 171
DB 126 LQVNSPSSGRISDSSGLGEPEFVENFNPYSGWGRPQDDGALPALTMTLTYRQLIQGK 185
QY 172 TSTATDVLWPLVRNDLSVAQYNNQGTGYDLWEVNGSFFPIAVQHRALVEGSAFAATVAG 231
DB 186 QSVASNLIMFPVANDLITVAQYNNHTGDLWEELIDSSFFTTVAQHRAMWEGSALQAALG 245
QY 232 SSGCWDSQADEILICYIQSFWTGSPFIANF--DSSRGKDANTLIGSIHTFPDPAACDS 289
DB 246 KPHAGYAVAVPEILICLLQSYNNESALISININNNRGSGIDINSVLTSHITPDPAACDS 305
QY 290 TFGQCSRRALANHEVUDFSRSTYTLNDGLSDSEAVAVGRPEPTYNNNGNWFICTLAAA 349
DB 306 TFGQCSRRALANHRYVUDFSRSTYGINAGLPGRAANVGRVAEDIYQGGNFWYLTATLAAA 365

QY 350 EQLYDALYQWQKGSLEVTDVSLDFPKALYSDAATCTYSSSSSTYSIYDAVKTPADGV 409
DB 366 EQLYDALYQWQKGSLEVTDVSLDFPKALYSDAATCTYSSSSSTYSIYDAVKTPADGV 425
QY 410 SIYETHAANSNGMSBEQYDSDGEQLSARPLTWSTYALTLANNRRNSVPAWSEETASAV 469
DB 426 FLVEKTPSPNGSLAEYDRNTGVPLSANDLTWSTYAAFLSTLQRLNIMPDSWGPSSANTV 485
QY 470 PGTCAATSAIGTYSSV 485
DB 486 PTTCSKTITITGYSAV 501

RESULT 37

AAW30155 ID AAW30155 standard; protein; 616 AA.

XX AAW30155;

XX 17-OCT-2003 (revised)

XX 25-MAR-2003 (revised)

XX 15-JAN-1998 (first entry)

XX Glucoamylase P.

XX glucoamylase P; horomonis resinae; debranching activity; enzyme;

XX dextrinase activity; Trichoderma reesei; starch granule; preservation;

XX hydrolysis; gelatinised starch; glucose syrup; straight linear dextrin;

XX saccharification; lignocellulosic material; sugar utilisation;

XX feed additive.

XX Amorphotheca resinae.

XX US5655585-A.

XX 09-SEP-1997.

XX 07-FEB-1995; 95US-00385370.

XX 03-SEP-1992; 92US-00937789.

XX 12-AUG-1993; 93US-00104853.

XX (ALKO-) ALKO-YHTIOT OY.

XX Jouteloki V, Vainio A, Fagerstrom R, Nevalainen H, Aho S;

XX Torkkeli T, Korhola M, Torkkeli H;

XX WPI; 1997-456802/42.

XX N-PSDB; AAT90830, AAT90831.

XX Hormoonis resinae glucoamylase P gene construct - for transforming

XX Trichoderma to produce recombinant glucoamylase P.

XX Claim 1; Col 58-60; 61pp; English.

XX This sequence represents the Hormoonis resinae glucoamylase P. The DNA
XX encoding this sequence is used in the composition of the invention and
XX are capable of being processed by a Trichoderma host cell. H. resinae
XX glucoamylase P has higher debranching and dextrinase activity than
XX conventional glucoamylase-pullulanase mixtures. T. reesei secretes
XX enzymes that are important for the degradation of complexes around and in
XX starch granules. The recombinant glucoamylase P enzyme produced by the
XX Trichoderma host cell is useful in applications requiring the hydrolysis
XX of gelatinised starch or the presence of a debranching activity. These
XX applications include starch analysis, the manufacture of glucose syrups,
XX production of straight linear dextrin for use in food, medicines and
XX cosmetics, and in the preparation of food fibers by the enzymatic
XX treatment of laundry hanks or bales. The enzyme can also be used as an
XX additive to laundry and dish washing detergents, in wood and textile
XX industry applications such as the preparation of plywood adhesives. It
XX can also be used in the saccharification of lignocellulosic materials,
XX the preservation of protein-containing animal or vegetable fodder, the

CC preparation of feedstuff from meat by-products and to improve utilisation
CC of sugar present in such feedstuff. When used as a feed additive for food
CC animals, the recombinant bacterial biomass may be added directly to the
CC animal's fodder. (Updated on 25-MAR-2003 to correct PF field.) (updated
CC on 17-OCT-2003 to standardise OS field)

XX Sequence 616 AA;

Query Match 47.7%; Score 1321; DB 2; Length 616;
Best Local Similarity 48.6%; Pred. No. 1.8e-97;
Matches 257; Conservative 88; Mismatches 150; Indels 34; Gaps 6;

QY 14 CTGLANVIS-----KRALDMSLNEMATVARTALINMGADAWGSGSIVVA 63
DB 8 CGALSLTSLGSLAIAATELKARDLSFISERAIALOGALNINGPBGSAVPGAGAFVVA 67
QY 64 SPSTNDPDIYPTWTRDSGLVLTVDLFRNGDSTLSLTENYISAQAIVOGISNPGD-L 122
DB 68 SPSKANPDYFTWSRDSALTKMIIDEFLIGNTLTQITIEQYTHAQAVIQVSNPSCGTEL 127
QY 123 SSGAGIGEPKFNVDFAVYSGWGRPORDPALATAMIGCGWLNDNGYTSATDITVPL 182
DB 128 PPGVGLGEPRFNVDPGRFNGPWRPORDPALALALMTYSNWLINQGFAPKTKIWP1 187
QY 183 VANDLSYVAQYMNQGYDLMEEVNCGSPFTIAVOHRAIVEGSAFATAVGSSCSWCSQAP 242
DB 188 INADLSYVQYMNQSGFDLMEERYASFTTIQNGHALVBNQALMDLQVTCGCC-QAP 246
QY 243 ELICYIQSTWTSFIANP--DSSRGKANTLLGSIHTPDPACDSTFQPCSPRALA 300
DB 247 EYLCEIQSTWTSFIANP--DSSRGKANTLLGSIHTPDPACDSTFQPCSPRALA 306
QY 301 NKEVVDSPRTYTLNDGSLDSBAVAVGRYPEDTYNNGNPFCTLAABEOLYDALYOMD 360
DB 307 NKEVVDSPRTYTLNDGSLDSBAVAVGRYPEDTYNNGNPFCTLAABEOLYDALYOMD 366
QY 361 KCGSLEVTVDLDFKALYSDAATGYSS--SSSTYSIYDAKTRADGFSIVETHAAS 418
DB 367 AHHVLTVDSTLAFDQIPEVTVREYKSGNANSPQAQIMDAVTAADSVVAIAEKYIIS 426
QY 419 NSGMSQYKSDGEOULSARDLTWSVYALLTANNRRNSVPASGETSASSVPCTCAATSA 478
DB 427 NSLSIQFNRDICTPLAIDLITWSVYAFITMSQRRAGQYPSWSGSHNALPPPTCSASST 486
QY 479 IGTYSVVYTWSPSIYATGTTTATPTGSGVTSKTTATASKSTT 527
DB 487 PGII-----TPATAAGAPNVTSSCQVSTFENINATT 517

RESULT 38

ABP96632 standard; protein; 579 AA.

XX ABP96632;

DT 02-JUN-2003 (first entry)

XX Rhizopus oryzae glucosylase protein sequence SEQ ID NO:49.

XX Self-processing plant; plant; processing enzyme; alpha-amylase; grain;
XX pullulanase; alpha-glucosidase; glucose isomerase; glucoamylase;
XX mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;
XX maltodextrin; ethanol; fermentation; beverage; enzyme.

XX Rhizopus oryzae.

XX WO2003018766-A2.

XX 06-MAR-2003.

XX 27-AUG-2002; 2002WO-US027129.

XX 27-AUG-2001; 2001US-0315281P.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.
PA Lanahan MB, Basu SS, Batie CJ, Chen W, Craig J, Kinkema M;

XX MPI; 2003-268420/26.

XX N-PSDB; ACC44574.

XX Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.

PT alpha-amylase, useful for producing plant to produce food products having

PS Claim 1; Page 110; 158pp; English.

CC The present invention describes polynucleotides which encode processing
CC enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose
CC isomerase, or glucoamylase) that are optimised for expression in plants.
CC The polynucleotides encode mesophilic, thermophilic or hyperthermophilic
CC processing enzymes, which are activated under suitable conditions to act
CC upon the desired substrate. Also described are self-processing transgenic
CC plants and plant parts, e.g. grain, which express one or more of these
CC enzymes and have an altered composition that facilitates plant and grain
CC processing. Also described is a method (M) for converting starch to
CC starch-derived products in a transformed plant part (TPP), by activating
CC the starch processing enzyme contained in it. Transgenic grain is useful
CC for preparing maltodextrin. A transformed plant (TP) can be used to
CC produce food products having improved taste and to produce fermentable
CC substrates for ethanol and fermented beverages. (M) eliminates the need
CC to mill or physically disrupt the integrity of plant parts prior to
CC recovery of starch-derived products. The present sequence represents
CC glucamylase, which is given in the exemplification of the present
CC invention

XX Sequence 579 AA;

Query Match 26.2%; Score 725; DB 6; Length 579;
Best Local Similarity 36.8%; Pred. No. 2e-49;
Matches 165; Conservative 71; Mismatches 170; Indels 42; Gaps 10;

QY 25 ATLDWSLNEMATVARTALINMGADAWGSGDGLVNASPSTNDPDIYPTWTRDSGLVL 84
DB 143 STISSWIKKQEGISPRAMRNINP-----PGSATGFIAMSLTAGDYYVAMWRDALTS 197
QY 85 KTLVDFR--NGDPSLTSTENYISAQAIVOGISNPGDLSGAGIGEPKFNVDFAVY 141
DB 198 NVIYVEYNTLTSKNTILVAKDYTFYSVKTOSTSVNC-----CLGEPKFNPDASGTT 251
QY 142 GSWGRPORDPALATAMIGFC-OWLLDNGYTSATDIWPLVRNDLSVYAOYMNQGYD 200
DB 252 GAMGRQNDGPAERATTFILFADSVLTORKDASVYGTILKPAIFKOLDVYVWVMSNGCFD 311
QY 201 LMBEYVNGSFFFTIAVOHRAIVEGSAFATAVGSS--CSWCSQAPETLCTLOSTWTSFLL 258
DB 312 LMBEYVNGVFTYLMWRKGLLGADFAKRNQDSTRASTYSSTASTANISSFWSSNNW 371
QY 259 ANPDS-----SRSGDANTL-----IGSIHTDPBAACDSTFQPCSPRALANHKEVDS 308
DB 372 IOVSQSVTGVSGSKGIDVSTLAAANGSV-----DDGFTTPESEKILATAVAVERDS 422
QY 309 FRSIYTLNDGLDSBAVAVGRYPEDTY-----YNGNPFCTLAABEOLYDALYOMDKOG 363
DB 423 FASIVYINKNLDSYNGNSIGRYPEDTYNNGNSQGSWFLAYTGVAELLYRAIKEMWING 482
QY 364 SLEVTVDLDFKALYSDAATG-TYSSSSTYSIYDAKTRADGFSIVETHAASNGSN 422
DB 483 GVTVSSISLPPFKKPDSSATSGKTYVGTSDNNILAQNTALADRFLSTVQJAHANNGL 542
QY 423 SEQYKSDGEOULSARDLTWSVYALLTAN 450
DB 543 ABEFRTTGLSTGARDLTWSHSLITAS 570

RESULT 39

```

AAP60359
ID AAP60359 standard; protein; 604 AA.
AC AAP60359;
XX
XX 23-MAY-1991 (first entry)
DT
DE Glucoamylase structural gene product.
XX
XX Protease; Rhizopus; amylase; alcohol.
XX
XX Rhizopus oryzae.
OS
XX
XX Key Location/Qualifiers
XX Peptide 1..25
XX FT /label= signal peptide
XX FT 26..604
XX FT Protein /label= Mature peptide
XX
XX EPI86066-A.
XX
XX 02-JUL-1986.
XX
XX 13-DEC-1985; 85EP-00115910.
XX
XX 15-DEC-1984; 84UP-00264964.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Aahihari T, Nakamura N, Tanaka Y, Shibano Y, Yoshizumi H;
XX WPI, 1986-170627/27.
XX DR N-PSDB; AAN60292.
XX
XX Gluco-amylase gene obtd. from rhizopus strain - useful in transforming
XX PT yeasts or bacteria for efficient prodn. of glucoamylase for starch
XX PT hydrolysis in alcohol prodn.
XX
XX Claim 3; Fig 1; 108pp; English.
XX
XX The gene may be used to form vectors capable of transforming an E.coli or
XX CC B subtilis expression sysem for the economical production of the enzyme.
XX CC Glucoamylase has a good enzymatic activity for hydrolysing raw starch
XX CC especially in the manufacture of alcohol, and is not degraded by protease
XX CC activity during production
XX
XX Sequence 604 AA;
SQ
Query Match 26.1%; Score 724; DB 1; Length 604;
Best Local Similarity 36.8%; Pred. No. 2.6e-49;
Matches 165; Conservative 71; Mismatches 170; Indels 42; Gaps 10;
QY 25 ATLDLWLSNEATVARTAIINIGADGAWVSGADSGIVVAPSTNDPPYFTWTRDGLVL 84
DB 168 STISSWTKKQEGISRFPMKLNINP-----PGSATGFPAASLSTGPPYTYAMTRDALTS 222
QY 85 KTLVDLFR--NGDTSLLSTIENYISAQAIVQGISNPSGDLSSGAGLGEPEKFNVDETAYT 141
DB 223 NVLIYEYNTLTLSGNKKTILNVLKDYVTFPSVKTQSTSYTCN-----CLGEPEKFPNDGSGYT 276
QY 142 GSNRGFRPDGPALRATMIGFG-OMLIDNGYTATDIWPLVRNDLSYVAQIWNQGTGYD 200
DB 277 GAWGRPQNDGPAAERATTFILFADSYLTQTQDASVYVTLKPAIFKDLVYVNVWWSNCGFD 336
QY 201 LMEEVNNGSPFTTAVOHRALVEGSAFATVAGSS--CSWCDGQAPEILCYLOSFWTGSFIL 258
DB 337 LMEEVNNGVHYTTLWVWKGILGADPFAKRGDSTRASTYSTASTIANKISSFWVSSNW 336
QY 259 ANPDS-----SRGKDANTL---LQSIHTFPDEAACDDSTFQPCSPRALANKEVVD 308
DB 397 IQVSQSVTGAVSKKGLDVLSTLANLGSV-----DGGFTPGESEKTLATVANVEDS 447
QY 309 FRSIYTLNDGLDSEAVAVGRYPEDTY-----YNGNFWPLCTLAAEQLYDALYQMDXG 363

```

```

DB 448 FASLYPINKNLPSYLNSIGRYPEDTYNGNGNSQGSNWFPLAVTGYALYTRAKKWIWG 507
QY 364 SLEETDVSIDPFKALYDAATG-TYSSSSSTVSSIYDAVKTADGVSIVETIAANGSM 422
DB 508 GVTYSSISLDPFKKPFSSSATSGKKYTVGTSDPFNNLALNADRLSLTYQLAHANNGL 567
QY 423 SEQYKSDGBQLSARDLTWSYAALLTAN 450
DB 568 AEEFDRRTGTGARDLTWSSHASLTITAS 595
RESULT 40
AAR77674
ID AAR77674 standard; protein; 624 AA.
XX
XX AAR77674;
XX
XX 25-MAR-2003 (revised)
XX DT 29-JUN-1996 (first entry)
XX
XX Glucoamylase from Arxula adeninivorans.
XX
XX glucoamylase; yeast; Saccharomyces cerevisiae; production; starch;
XX KM glucose; degradation; carbon source.
XX
XX Arxula adeninivorans.
XX
XX Key Location/Qualifiers
XX FT Peptide 1..16
XX FT /label= signal_peptide
XX FT Protein /note="for secretion of glucoamylase"
XX FT /label= mature_peptide
XX FT Modified-site 54..56
XX FT /note="potential N-glycosylation site"
XX FT Modified-site 70..72
XX FT /note="potential N-glycosylation site"
XX FT Modified-site 98..100
XX FT /note="potential N-glycosylation site"
XX FT Modified-site 111..113
XX FT /note="potential N-glycosylation site"
XX FT Modified-site 168..170
XX FT /note="potential N-glycosylation site"
XX FT Modified-site 267..269
XX FT /note="potential N-glycosylation site"
XX FT Modified-site 333..335
XX FT /note="potential N-glycosylation site"
XX FT Modified-site 460..462
XX FT /note="potential N-glycosylation site"
XX PN DE4425058-AI.
XX
XX 18-JAN-1996.
XX
XX 15-JUL-1994; 94DE-04425058.
XX
XX 15-JUL-1994; 94DE-04425058.
XX
XX (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.
XX
XX Kunze G, Bui M, Kunze I, Foerster S;
XX
XX WPI, 1996-069579/08.
XX DR N-PSDB; AAT08701.
XX
XX Prodn. of heat stable glucoamylase in yeast able to utilise starch - by
XX PT transformation with glucoamylase gene including secretion sequence from
XX PT Arxula adeninivorans.
XX
XX Claim 1; Fig 7; 21pp; German.
XX
XX The present sequence is a heat stable glucoamylase derived from Arxula
CC

```

CC adeninivorans. Yeast, e.g. *Saccharomyces cerevisiae*, can be transformed
CC with the DNA (AA08701) for the production of the enzyme which is able to
CC used starch as its carbon source. The enzyme has the N-terminal,
CC secretion sequence shown in AA07675. The enzyme is useful for
CC degradation of starch, giving glucose as a final product. In yeast the
CC DNA encoding the enzyme can be easily manipulated to produce the protein
CC in high yield, which is otherwise difficult to do in *A. adeninivorans*.
CC (updated on 25-MAR-2003 to correct PN field.)

SQ Sequence 624 AA;

Query Match 23.2%; Score 641.5; DB 2; Length 624;

Best Local Similarity 32.8%; Pred. No. 1.2e-42;
Matches 154; Conservative 90; Mismatches 172; Indels 53; Gaps 13;

QY 27 LSWL--SNEAT--VARTALNNIGADGAMVSGADSGIVASSTNDPDYFYTWTRDSGL 82
DB ITNWLKPSNDSTGKLSFLFNNIN----IPGAGTVIAAQSYSEPDYATWVADSL 214
QY 83 VLKTLVDLFRNGDTSLLSTIENYISAQAIVOGISNPSGSLSGAGLGEKPFNVDETAATG 142
DB 215 VMDVYNRLYSSAKSEKQLYEKILFYAKAG-AQEQNDPTAISGMEKFTYIANTAFITG 273
QY 143 SWGRPDQDPALPATAMIGFGWLLDNG-----YTSTATDIWPLVRNDLSYVAQ 192
DB 274 SWGRPDNDGPATRAITLIEFANAYLANGSGOPTVREQLYDSKYPQVAP-IKKDLQFVAS 332
QY 193 YNNGCYDLMEEVNGSFFTIAYOHALYEGSAFAFAVSS--CSWCDQAPAILCYLOS 250
DB 333 NMSSPFDLMEEBEESAHFYTRLVQRKALLGADFADMDKDELSDKLTQASLSDTLPB 392
QY 251 FWTSFILLNFD-----SSRSKGDANTLLGSIHTDPBAACDDSTFGQSPALANHKE 304
DB 393 FPDASRQLLYEYGPVLRKRYKYKDISVLGVWNGH-----ANDNVFSTINDQILATAQ 447
QY 305 VVDSFRSITLNDGLSDS-----AVAGRYPEPTY-----YNGNPMFLCTLAAEQLYDA 355
DB 448 VSTSLFDVYKVNANTSDSESGKPLGIPVGRYPEDVDGVTGSGNPMYLTMMAAEFLYS 507
QY 356 LYQMDKQGLEVTVDVSLDFPKALYSDA---ATCTYSSSSSYSIYDAVKTADGVSTIV 412
DB 508 VQEFEDAGSIITISDTSLPFWKYFASSVDHAKAKYKNDQSPKTSLSKSLTGMDADAFMRA 567
QY 413 EHTAASNGSSEYOYDKSDGQSLARDLTWSYALTLTA-----NNRN 454
DB 568 KHTTSSGMSSEFNRITTEPRGAKDLTWSYSLLSAFAREELRNQKN 616

RESULT 41
AAP70572
ID AAP70572 standard; protein; 497 AA.

XX AAP70572;
AC
XX
XX 24-OCT-2003 (revised)
DT 03-MAY-1991 (first entry)
XX
DE Glucoamylase gene product from plasmid pSE glu 1.
XX
XX Amylase.
XX
XX
XX Saccharomycopsis fibuligera; HUT7212.
OS
XX JP62104576-A.
PN
XX
XX 15-MAY-1987.
PD
XX
XX 31-OCT-1985; 85JP-00244892.
PF
XX
XX 31-OCT-1985; 85JP-00244892.
PR
XX
XX (FUKUI) FUKUI S.
PA
XX

DR WPI; 1987-173694/25.
DR N-PSDB; AAN70917.

XX
XX Amylase prodn. - comprises culturing microorganism transformed with
PT vector deoxyribonucleic acid, accumulating and collecting amylase.
XX
XX Disclousure; Fig 1; 14pp; Japanese.
PS
XX

CC The product is an amylase biosynthetic enzyme encoded by a plasmid which
CC may be used to transform an E.coli expression system for the stable
CC production of amylase, useful in ethanol fermentation. (Updated on 24-OCT
CC -2003 to standardise OS field)
CC
XX
XX

SQ Sequence 497 AA;

Query Match 22.2%; Score 615.5; DB 1; Length 497;

Best Local Similarity 34.1%; Pred. No. 1.1e-40;
Matches 158; Conservative 73; Mismatches 156; Indels 77; Gaps 16;

QY 24 RATLDSWLSNEATVARTALNNIGADGAMVSGADSGIVASPSNDPDYFYTWTRDSGL- 82
DB 42 RDLERFLDKQKGVSLYLQNIATPEGPFNNGVRETIVASISTSPDYQWTRDSALT 101
QY 83 VLKTLVDLFRNGDTSLLSTIENYISAQAIVOGISNPSG--DLSSGAGLGEKPFNVDE 137
DB 102 FLTVLSELD--NFEVTLAKAVEYINTSYNLTQRTSNPSGFFDENHKGGLGEKPFNTDG 159
QY 138 TAYTSWGRPDQDPALPATAM-----IGFGWML-----DNCYTSTANDIWPPLVR 184
DB 160 SAYTAGMGRPDQDPALPATAMSRVLYNDVNSLNEKVLTLDSGDIINFST-EDYKNIK 218
QY 185 NDLSYVAQYWNQGYDLMEEVNGSFFTIAYOHALV-----EGSAFAFAVSSCS 235
DB 219 PLEEVYIGVWDSITGPDLMENGRHFTSLVQKALAYAVDLAKSRDDDDFANTLSTTS 278
QY 236 WCDQAPAILCYLOSFWTGS-----FILLNFD-----SSRSKGDANTLLGSIHTPD 281
DB 279 -----TLSEYLSGSDGFAVNTDVNHLVENDLLQNSRQGLDQATYIGPLTHD 327
QY 282 PEACDDSTFGQSPALANHKEVVDSPSITLNDGLSDSEAVAVGRPEPTY-----Y 336
DB 328 -IGESSSTFPDVNDNEVYLOSYYLLLEDNDRYSVNSAY--SAGAAIGRYPEPDYNGDSS 384
QY 337 NGNPMFLCTLAAEQLYDALYQMDKQGLEVT--DVSLDF-----KALYSDAAT 384
DB 385 EGNPMFLATAYAAQPYKALYD-AKSASNDITINKINYDFNKYIVDLSTINSAYSSDS 443
QY 385 GTYSSSSSTYSIVDAVKTADGFVSIYETHAASNGSMEYOYDK 428
DB 444 VTIKSGSDEFNTVADNLVTFGDSFLQVILDHINDGSLNEQLNR 487

RESULT 42
AAP60723
ID AAP60723 standard; protein; 918 AA.

XX AAP60723;
AC
XX
XX 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 08-JUN-1991 (first entry)
XX
DE Sequence of extracellular amylase-1,4-glucosidase (AMG) precursor.
XX
XX Precursor polypeptide; secretion vector; enzyme.
XX
XX Saccharomycetes cerevisiae.
XX
XX
XX
XX
XX Key Location/Qualifiers
FT Peptide 42. 62
FT /label= signal
FT /note= "claimed"
FT Protein 63. 918

DB 287 YFNA 290

RESULT 44
ADCO1413
ID ADCO1413 standard; protein; 1588 AA.

XX AC ADCO1413;
XX DT 04-DEC-2003 (first entry)

DE Enterohemorrhagic E. coli O157:H7-specific protein SEQ ID NO: 1459.
XX enterohemorrhagic; anti-bacterial.
KM Escherichia coli; O157:H7.
XX JP2002355074-A.
XX PN 10-DEC-2002.
XX PD 24-JAN-2002; 2002JP-00015959.
XX PF 24-JAN-2001; 2001JP-00112010.
XX PR (UYTS-) UNIV TSUKUBA.
XX PA WPI; 2003-451640/43.
XX DR WPI; 2003-451640/43.

XX PT Enterohemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule
XX and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX PS Claim 3; SEQ ID NO 1459; 2067pp; Japanese.

CC The invention relates to a novel enterohemorrhagic Escherichia coli
CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC has anti-bacterial activity. The polypeptide can be used in detection
CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC genome of Enterohemorrhagic E. coli O157:H7 was determined. The present
CC sequence represents an E. coli O157:H7-specific polypeptide of the
CC invention.

SQ Sequence 1588 AA;

Query Match 6.5%; Score 181; DB 7; Length 1588;
Best Local Similarity 21.7%; Pred. No. 7.2e-05; Indels 178; Gaps 30;
Matches 136; Conservative 95; Mismatches 217;

QY 15 TGLANTISKRAITDLSWLSNEATVARTAILNIGADG-AMVSGADSGIVASPTDNDPDPF 73
DB 576 TINIANTTSINATITNINISN-----LLETYNLGEDALKM--DKDNGVFTAAHGETTTSKI 628
QY 74 YWTRD-----SGLVLTIVDLFENGDTSLSTIENYISAQAIVQGISNPSGDL 122
DB 629 -FNVKGDGLTTSSTDAVNGSQLKTNDVAVNTNTNATNTN-----ISNLETIV 677

QY 123 SSAGAGGEP--KENVDEIATVSGMRPQRGPRALRATAMGFGOMLIDNGYTATDIW 180
DB 678 TN---GGEALAKWKDKNGVFTAAHG---NNTASKITN-----ILDGTVATISSDAIN 723
QY 181 PLVRNRL-SYVAQYM-----NOTG-----YDLMEEVNSSFPTIAVQHRALVESGAPA 227
DB 724 GSQLYLLSSNIAIYFGNMSVNTDGVFTGPTYKIGF---TNYVNGDALAAI--NSSFS 777

QY 228 TAVSSSCSWCDQAPELLCLYQSPWTGSPF--LANPDSRSRGA--NTLLGSIHTF--- 280
DB 778 TSLGDALLW-DATAGKFSAGKGTNGDASVITVDAGBISDSSSDAVAVNGSLHGVSSVVD 836
QY 281 ----DEACDDSTFQPCSPRALANKEVVDSEFRST-YTINDGLSDSEAVAVGRYPEDTX 335
DB 837 ALGGGAEVNADGTTTAPTYTIANADYDNDVADALNADITLDDAL-----LMDADAG 887
QY 336 YNGNPWFLLCTLAAEQLYDALYQMDKQSLVTVDSLDFKALYSDAATG----- 385

DB 888 ENG-----AFSAHGKDKTASV-ITNVANGAISAASDAINGSQLYTTNKYI 933
QY 386 -----TYSSSSSTYSIDAVTF-----A 405

DB 934 ADALGGDAEVNADGTTTAPTYTIANAEYNNVGDALDALLMDDETANGAGAVNASH 993
QY 406 DGFVSIVETHAASNGSMSE-QYDKSPDEQLSA-----RDLTWSYAL 446
DB 994 DKAASITN--VANGSISEDSTDAVNGSQLMNTNMIEQNTQIINQAGTDAITYOENG 1051

QY 447 LTAANNRNSVWPASWGETSASSVPGTCAATSAIGTSSVTVTSWPSIVATGTTTATPT 506
DB 1052 AGINIVRTNDDGLAFNDASAQGVGAT-----AIG-YNVAKGDSVAIGGYSVDVTGI 1105

QY 507 GSGSVTSTSKTTRTATSKTSTTTRSGM 532
DB 1106 ALGSSSVSSSRVIKGRSDTSITENG 1131

RESULT 45
AAR22343
ID AAR22343 standard; protein; 32 AA.

XX AC AAR22343;
XX DT 25-MAR-2003 (revised)
XX DT 08-MAY-1992 (first entry)

XX DE His(184) mutation in A. awamori glucoamylase region 4.
XX 1,4-alpha-D-glucan glucosylhydrolase; EC 3.2.1.3; starch hydrolase.
XX OS Aspergillus awamori.
XX PN WO9200381-A.
XX PD 09-JAN-1992.
XX PF 29-JUN-1990; 90US-00546511.
XX PR 29-JUN-1990; 90US-00546511.
XX PA (NOVO) NOVO-NORDISK AS.
XX PI Svensson KB, Sierks MR;
XX N-PSDB; AAQ20804.
XX WPI; 1992-041562/05.

PT New mutated glucoamylase enzymes - having aminoacid changes for increased
PT selectivity for alpha-(1,4)-glucoside bonds in starch hydrolysis.
PS Claim 18; Fig 2; 42pp; English.

CC This sequence represents amino acids 165 to 196 of the A. awamori
CC glucoamylase enzyme, but with Ser 184 mutated to His. The mutation is in
CC the region corresponding to Region 4 (residues 172-184) of the A. niger
CC enzyme. The mutation confers increased selectivity for maltose hydrolysis
CC on the enzyme expressed in Saccharomyces cerevisiae. (Updated on 25-MAR-
CC 2003 to correct PA field.)

SQ Sequence 32 AA;

Query Match 6.2%; Score 172; DB 2; Length 32;
Best Local Similarity 96.9%; Pred. No. 9.9e-07; Indels 0; Gaps 0;
Matches 31; Conservative 0; Mismatches 1;

QY 189 YVAQYNNQGTGYLMEEVNSSFPTIAVQHRAL 220
DB 1 YVAQYNNQGTGYLMEEVNSSFPTIAVQHRAL 32

Search completed: June 28, 2004, 07:43:20
Job time : 65 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:40:37 ; Search time 46 Seconds
(without alignments)
3662.755 Million cell updates/sec

Title: US-10-038-723-2

Perfect score: 2771

Sequence: 1 MSFRSLALSGLVCTGLANV.....SKTTATASKTSTTRSGMSL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 100 summaries

Database :

SPTREMBL_25:*

- 1: sp archaea:*
- 2: sp bacteria:*
- 3: sp fungi:*
- 4: sp human:*
- 5: sp invertebrate:*
- 6: sp mammal:*
- 7: sp mhc:*
- 8: sp organelle:*
- 9: sp phage:*
- 10: sp plant:*
- 11: sp rodent:*
- 12: sp virus:*
- 13: sp vertebrate:*
- 14: sp unclassified:*
- 15: sp_evirus:*
- 16: sp_bacterioph:*
- 17: sp_archaeoph:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2669.5	96.3	639	3	Q870G8
2	2564.5	92.5	639	3	Q12537
3	1689	61.0	618	3	Q9C1V4
4	1518	54.8	493	3	Q59846
5	1400	50.5	620	3	Q12523
6	1394	50.3	579	3	Q12596
7	1286	46.4	571	3	Q9P4C5
8	784	28.3	609	3	Q8J0P8
9	728	26.3	579	3	Q727X9
10	674	24.3	515	3	Q8TFE5
11	187	6.7	589	17	Q96210
12	185.5	6.7	636	17	Q9H188
13	181	6.5	1588	16	Q8XDG4
14	175	6.3	1236	3	Q9C105
15	174.5	6.3	659	17	Q97BM7
16	174.5	6.3	990	16	Q83J22

17	173.5	6.3	1713	3	Q8TGE1	Q8TGE1 saccharomyc
18	172	6.2	622	17	Q97ZD0	Q97ZD0 sulfolobus
19	171	6.2	3360	16	Q8BXB6	Q8BXB6 lactobacill
20	170.5	6.2	1275	5	Q76602	Q76602 caenorhabdi
21	169.5	6.1	190	3	Q07070	Q07070 saccharomyc
22	168	6.1	4106	16	Q8XQP2	Q8XQP2 raietonia s
23	164	5.9	2016	5	Q9BIT0	Q9BIT0 plectreureys
24	163.5	5.9	457	5	Q86AK1	Q86AK1 dictyosteli
25	163	5.9	1333	16	Q8PD18	Q8PD18 xanthomonas
26	161	5.8	1663	16	Q8CH86	Q8CH86 lactococcus
27	161	5.8	2283	2	Q8WQ99	Q8WQ99 staphylococ
28	160.5	5.8	916	5	Q7YZ10	Q7YZ10 monosiga br
29	160	5.8	615	17	Q973T2	Q973T2 sulfolobus
30	160	5.8	888	5	Q25336	Q25336 leishmania
31	159	5.7	2271	16	Q99QV4	Q99QV4 staphylococ
32	158	5.7	1283	3	Q9URU4	Q9URU4 schizosacch
33	156.5	5.6	2232	5	Q8IFX6	Q8IFX6 caenorhabdi
34	156	5.6	1306	2	Q93N36	Q93N36 pantoea ana
35	154.5	5.6	2275	16	Q8NUJ3	Q8NUJ3 staphylococ
36	153	5.5	1131	3	Q74851	Q74851 schizosacch
37	153	5.5	2271	16	Q8E473	Q8E473 streptococc
38	152.5	5.5	2117	3	Q9EUI1	Q9EUI1 neurospora
39	152	5.5	2014	16	Q7U3X4	Q7U3X4 synecococc
40	150	5.4	1383	3	Q874K9	Q874K9 candida alb
41	149	5.4	967	3	Q08294	Q08294 saccharomyc
42	149	5.4	1001	3	Q05164	Q05164 saccharomyc
43	147	5.3	2230	16	Q7U7J7	Q7U7J7 synecococc
44	146	5.3	612	17	Q97VY0	Q97VY0 sulfolobus
45	145	5.2	860	16	Q8ET89	Q8ET89 lactobacill
46	144.5	5.2	614	16	Q9CHH3	Q9CHH3 lactococcus
47	144	5.2	948	3	Q74346	Q74346 schizosacch
48	143.5	5.2	881	2	Q9AF09	Q9AF09 cellvibrrio
49	143	5.2	1995	17	Q8T164	Q8T164 methanosarc
50	142	5.1	536	3	Q9UVS8	Q9UVS8 aspergillus
51	142	5.1	854	16	Q89GQ1	Q89GQ1 bradyrhizob
52	142	5.1	1672	16	Q8Y366	Q8Y366 raietonia s
53	141.5	5.1	1441	16	Q9CF11	Q9CF11 lactococcus
54	141.5	5.1	3145	16	Q98MG7	Q98MG7 rhizobium l
55	140.5	5.1	1055	16	Q839R5	Q839R5 enterococcu
56	140.5	5.1	1952	5	Q8WVW6	Q8WVW6 naegleria g
57	140.5	5.1	4776	16	Q97P71	Q97P71 streptococ
58	140.5	5.1	5188	16	Q8X4H5	Q8X4H5 escherichia
59	140	5.1	1802	3	Q04051	Q04051 saccharomyc
60	140	5.1	1994	16	Q8G521	Q8G521 bifidobacte
61	139.5	5.0	2522	16	Q8EKA6	Q8EKA6 shewanella
62	139.5	5.0	4654	16	Q8D418	Q8D418 vibrio vuln
63	139	5.0	402	10	Q23054	Q23054 arabidopsis
64	139	5.0	744	3	Q8TFG9	Q8TFG9 schizosacch
65	139	5.0	970	16	Q8DY17	Q8DY17 streptococ
66	139	5.0	1203	5	Q9N5K0	Q9N5K0 caenorhabdi
67	139	5.0	22152	4	Q8W417	Q8W417 homo sapien
68	138.5	5.0	600	5	Q86B01	Q86B01 dictyosteli
69	138.5	5.0	1589	5	Q46097	Q46097 drosophila
70	138.5	5.0	2310	16	Q8CMU7	Q8CMU7 staphylococ
71	138.5	5.0	5291	16	Q8X2T1	Q8X2T1 escherichia
72	138	5.0	316	12	Q8VOM4	Q8VOM4 equine herp
73	138	5.0	332	12	Q8VOM5	Q8VOM5 equine herp
74	138	5.0	337	12	Q8VOM1	Q8VOM1 equine herp
75	138	5.0	342	12	Q8VOM8	Q8VOM8 equine herp
76	138	5.0	356	12	Q8VOM7	Q8VOM7 equine herp
77	138	5.0	357	12	Q8VOM2	Q8VOM2 equine herp
78	138	5.0	372	12	Q8VOM3	Q8VOM3 equine herp
79	138	5.0	374	12	Q8VOM6	Q8VOM6 equine herp
80	138	5.0	389	12	Q8VOM0	Q8VOM0 equine herp
81	138	5.0	867	12	Q8VOM5	Q8VOM5 equine herp
82	138	5.0	867	12	Q8VOM5	Q8VOM5 equine herp
83	138	5.0	1347	3	Q87410	Q87410 candida alb
84	138	5.0	1778	16	Q8RCB2	Q8RCB2 schizosacch
85	137.5	5.0	481	5	Q9XUP4	Q9XUP4 caenorhabdi
86	137.5	5.0	625	3	Q74922	Q74922 schizosacch
87	137	4.9	354	16	Q8ET56	Q8ET56 lactobacill
88	136.5	4.9	1265	16	Q8EV72	Q8EV72 mycoplasma
89	136	4.9	314	16	Q8ET57	Q8ET57 lactobacill

90 136 4.9 866 12 039781 039781 equine herp
91 136 4.9 1180 16 09CH4 09CH4 lactococcus
92 136 4.9 3072 2 0939MS 0939MS streptococ
93 135.5 4.9 496 16 088VA6 088VA6 lactobacill
94 135.5 4.9 658 17 08PY7 08PY7 methanosarc
95 135 4.9 699 2 048494 048494 kurtzia zop
96 135 4.9 3624 16 082411 082411 salmonella
97 134.5 4.9 651 2 09KMR2 09KMR2 thermactin
98 134.5 4.9 930 2 09RFK5 09RFK5 caldiabacill
99 134.5 4.9 1014 5 0962B4 0962B4 helicoverpa
100 134 4.8 498 5 09GOU3 09GOU3 theileria p

ALIGNMENTS

RESULT 1
0870G8 PRELIMINARY; PRT; 639 AA.
ID 0870G8
AC 0870G8
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Glucoamylase
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eucotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=721;
RX MEDLINE=95066018; PubMed=7975554;
RA Zhong L., Tang G., Yang K.;
RT "Isolation and sequencing of glucoamylase gene from a glucoamylase
RT over producing strain."
RL Wei Sheng Wu Xue Bao 34:184-190(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=721;
RX
RA Zhong L., Qiao D., Tang G., Yang K.;
RT "Cloning, sequencing and comparison of the 5' flanking regions of glaA
RT gene from high and low glucoamylase-producing strains of Aspergillus
RT niger."
RL Wei Sheng Wu Xue Bao 36:181-186(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=721;
RA Zhong L., Qiao D., Tang G., Yang K.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY250996; AAF04499.1; -
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
DR GO; GO:0005976; P:polysaccharide metabolism; IEA.
DR InterPro; IPR002044; CBD 4.
DR InterPro; IPR008291; Glu-a-glcscd_SBD.
DR InterPro; IPR000165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_trans_6np.
DR Pfam; PF00686; CBM_20; 1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PRINTS; PR00736; GLHYDRLASE15.
DR ProDom; PD001566; CBD_4; 1.
DR PIRSF; PIRSF001031; Glu-a-glcscd_SBD; 1.
DR PIRSF; PIRSF001031; Glu-a-glcscd_SBD; 1.
SQ SEQUENCE 639 AA; 68308 MW; 025AD856B6542B69 CRC64;

Query Match 96.3%; Score 2669.5; DB 3; Length 639;
Best Local Similarity 98.1%; Pred. No. 3.9e-162;
Matches 521; Conservative 1; Mismatches 6; Indels 3; Gaps 3;

QY 1 MSFRLSLALSGVLTGLANYSKRATLDSWLSNEATVARTALININIGADGAWYSGADSGI 60
Db 1 MSFRLSLALSGVLTGLANYSKRATLDSWLSNEATVARTALININIGADGAWYSGADSGI 60
QY 61 VVASPSTDNPDYFTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIIVGGINPSG 120

Db 61 VVASPSTDNPDYFTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIIVGGINPSG 120
QY 121 DLSSGAGLGEPEKNVDEIATVTSWGRPORBGPALRATAMIGFQWMLDNGYSTANDIYW 180
Db 121 DLSSGAGLGEPEKNVDEIATVTSWGRPORBGPALRATAMIGFQWMLDNGYSTANDIYW 180
QY 181 PLVRNDLSVYAQYWNQTDYDLEEVNQGSEFTTAVOHRALVEGSAFATAVGSSCWCSDQ 240
Db 181 PLVRNDLSVYAQYWNQTDYDLEEVNQGSEFTTAVOHRALVEGSAFATAVGSSCWCSDQ 239
QY 241 APEILCYLQSFMTGSLFLANFDSRSRQKANTLLGSIHFPDEAQCDDSTFQCSFRL 239
Db 240 APEILCYLQSFMTGSLFLANFDSRSRQKANTLLGSIHFPDEAQCDDSTFQCSFRL 239
QY 300 ANHKEVVDSEFRSIYTTNDGLSDSEAAVAGRYPEDTYNGNPMFLCTLAABEQYDALYQW 359
Db 300 ANHKEVVDSEFRSIYTTNDGLSDSEAAVAGRYPEDTYNGNPMFLCTLAABEQYDALYQW 359
QY 360 DKQGSLEVTDVSLDFPKALYSDAATGYSSSSSTYSYDVAVCTPADGPIVYETHAASN 419
Db 360 DKQGSLEVTDVSLDFPKALYSD-ATGYSSSSSTYSYDVAVCTPADGPIVYETHAASN 418
QY 420 GSWSEQYDKSDGQSLARDLTWSYALALTANNRNSVVPASWGETSASVPGCAATSAT 479
Db 419 GSWSEQYDKSDGQSLARDLTWSYALALTANNRNSVVPASWGETSASVPGCAATSAT 478
QY 480 GTYSVTVTSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTTSS 530
Db 479 GTYSVTVTSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTTSS 529

RESULT 2
012537 PRELIMINARY; PRT; 639 AA.
ID 012537
AC 012537
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Glucoamylase precursor (EC 3.2.1.3).
GN GLA.
OS Aspergillus awamori.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eucotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=105351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X-100;
RA Digne I., Pechexonov V.T., Bulat S.A., Pirov L.M.;
RT "A glucoamylase gene from Aspergillus awamori X-100: structure,
RT allocation, and gene phylogeny."
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59303; AAB02927.1; -
DR HSSP; P04064; 1GAT.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005976; P:polysaccharide metabolism; IEA.
DR HSSP; P04064; 1GAT.
DR InterPro; IPR002048; EF-Hand.
DR InterPro; IPR008291; Glu-a-glcscd_SBD.
DR InterPro; IPR000165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_trans_6np.
DR Pfam; PF00686; CBM_20; 1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PRINTS; PR00736; GLHYDRLASE15.
DR ProDom; PD001566; CBD_4; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00820; GLUCOAMYLASE; 1.
DR PIRSF; PIRSF001031; Glu-a-glcscd_SBD; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 24
FT CHAIN 25 639 GLUCOAMYLASE.

SQ SEQUENCE 639 AA; 68277 MW; 6F93D0637D174ACB CRC64;
 Query Match 92.5%; Score 2564.5; DB 3; Length 639;
 Best Local Similarity 92.3%; Pred. No. 2e-155; Indels 11; Gaps 3;
 Matches 494; Conservative 16; Mismatches 14;
 QY 1 MSFRSLALSGVCTGLANVYISKRATLDSWLSNEATVARTALINNGADGAWYSGADSGI 60
 DB 1 MSFRSLALSGVCTGLANVYISKRATLDSWLSNEATVARTALINNGADGAWYSGADSGI 60
 QY 61 VVASPTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTENYISAQAIIVGQISNPSG 120
 DB 61 VVASPTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTENYISAQAIIVGQISNPSG 120
 QY 121 DLSGGIGEPKRVNDEATYTGSGWRPQRDPALRTAMIGFQWLLDNGYSTANDIYW 180
 DB 121 DLSGGIGEPKRVNDEATYTGSGWRPQRDPALRTAMIGFQWLLDNGYSTANDIYW 180
 QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNCSFPTIAYQHRALVEGSAFATVAGSSCNCDSQ 240
 DB 180 PLVRNDLSYVAQYWNQGYDLMEEVNCSFPTIAYQHRALVEGSAFATVAGSSCNCDSQ 239
 QY 241 APEILCYLOSFT---WTGSFIILNPDSSRSKGDANTLGSHTFPDEAACDDSTFQPCS 295
 DB 240 APEILCYLOSFT---WTGSFIILNPDSSRSKGDANTLGSHTFPDEAACDDSTFQPCS 294
 QY 296 PRLANHKEVVDSPRSTYTLNDLSSEAVAVGRYPEDTYNNGNPNWFLCTLAAEOLYDA 355
 DB 295 PRLANHKEVVDSPRSTYTLNDLSSEAVAVGRYPEDTYNNGNPNWFLCTLAAEOLYDA 354
 QY 356 LYOMDKQSLLEVTVSLDPEFKALYSDAATGTYSSSSSTYSIVDAVTFADGFSIVETH 415
 DB 355 LYOMDKQSLLEVTVSLDPEFKALYSDAATGTYSSSSSTYSIVSAKTFADGFSIVETH 414
 QY 416 AASNGSSEQYDKSDDEQLSARDLTVSYALLTANNRRNSVPAWGETSASSVPGTCAA 475
 DB 415 AASNGSSEQYDKSDDEQLSARDLTVSYALLTANNRRNSVPAWGETSASSVPGTCAA 474
 QY 476 TSAIGYSSVTWSPSIVATGTTTATPTGSGSVTSTKTTATASKSTTTTS 530
 DB 475 TSAIGYSSVTWSPSIVATGTTTATPTGSGSVTSTKTTATASKSTTTTS 529
 RESULT 3
 Q9C1V4 PRELIMINARY; PRT; 618 AA.
 AC Q9C1V4; 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Glucosylase precursor (EC 3.2.1.3).
 GN GA.
 OS Talaromyces emersonii.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; Talaromyces.
 OX NCBI_TaxID=68825;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Nielsen B.R., Lehmbeck J., Frandsen T.P.;
 RT "Cloning, heterologous expression, and enzymatic characterization of a
 thermostable glucosylase from Talaromyces emersonii";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ304803; CAC28076.1; -.
 DR HSSP; P04064; IGAT.
 DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
 DR GO; GO:0005976; P:polysaccharide metabolism; IEA.
 DR InterPro; IPR008291; Glu-a-glcSD_SBD.
 DR InterPro; IPR000165; Glyco_hydro_15.
 DR InterPro; IPR008928; Glyco_trans_6np.
 DR Pfam; PF00686; CBM_20; 1.
 DR Pfam; PF00723; Glyco_hydro_15; 1.

DR PRINTS; PR00736; GLHYDRASE15.
 DR ProDom; PD001568; CBD_4; 1.
 DR PIRSF; PIRSF001031; Glu-a-glcSD_SBD; 1.
 KM Glycosidase; Hydrolase; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 28 618 GLUCOSYLASE.
 SQ SEQUENCE 618 AA; 65429 MW; C48A034A2C065B2 CRC64;
 Query Match 61.0%; Score 1689; DB 3; Length 618;
 Best Local Similarity 61.2%; Pred. No. 1.4e-99;
 Matches 322; Conservative 75; Mismatches 111; Indels 18; Gaps 5;
 QY 7 LALSGVCTGLANV-ISKRT--LDSWLSNEATVARTALINNGADGAWYSGADGIIVA 63
 DB 9 LCLIGLTPAFAFAAPAAATAGSLDSFLATETPINALQVILNNGPAGADVAGSAGIIVA 68
 QY 64 SPSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTENYISAQAIIVGQISNPSGDL 123
 DB 69 SPSTNDPDYFYWTWTRDAALTAQYLVAFIAGNDLEQITIQYISAQAKYQTIENPSGDL 128
 QY 124 SGAGIGEPKRVNDEATYTGSGWRPQRDPALRTAMIGFQWLLDNGYSTANDIYWPLV 183
 DB 129 TG-GLGEPKRVNDEATYTGSGWRPQRDPALRTALIAVANYLIDNGEASTADEIIPV 187
 QY 184 RNDLSYVAQYWNQGYDLMEEVNCSFPTIAYQHRALVEGSAFATVAGSSCNCDSQAPE 243
 DB 188 QNDLSYITQYWNSTDLMEEVNCSFPTIAYQHRALVEGSAFATVAGSSCNCDSQAPQ 247
 QY 244 ILCYLOSFTWGSFIILNPDSSRSKGDANTLGSHTFPDEAACDDSTFQPCSAPALANH 302
 DB 248 VLCFLOSFTWGSFIILNPDSSRSKGDANTLGSHTFPDEAACDDSTFQPCSAPALANH 307
 QY 303 KEVVDSPRSTYTLNDLSSEAVAVGRYPEDTYNNGNPNWFLCTLAAEOLYDAIYQWKI 362
 DB 308 KVTVDSPRSTYTLNDSIAAGSAVAVGRYPEDTYNNGNPNWFLCTLAAEOLYDAIYQWKI 367
 QY 363 GLELVTVSLDPEFKALYSDAATGTYSSSSSTYSIVDAVTFADGFSIVETHAASNGSM 422
 DB 368 GSISITVDLSPFQDIPSAAVTIVSGTTFDIIISAVQTYDYLSTIVKTPEDGSL 427
 QY 423 SEQYDKSDDEQLSARDLTVSYALLTANNRRNSVPAWGETSASSVPGTCAATSAIGTY 482
 DB 428 TEQFSRTDGPPLASALTVSYALLTASARQGVPAWGESASSVPAWCAATSAIGTY 487
 QY 483 SSVTWSPSIVATGTTTATPTGSGSVTSTKTTATASKSTTTT 528
 DB 488 STATNTWPS-----SGSGSSTTSSAPCTPTPSAVT 520
 RESULT 4
 O59846 PRELIMINARY; PRT; 493 AA.
 ID O59846
 AC O59846; 01-AUG-1998 (TEMBLrel. 07, Created)
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Glucosylase.
 GN Aspergillus oryzae.
 OS Aspergillus oryzae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5062;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=O-1013;
 RX MEDLINE=98172744; PubMed=9511753;
 RA Hata Y., Ishida H., Ichikawa E., Kawato A., Suginami K., Imaiye S.;
 RT "Nucleotide sequence of an alternative glucosylase-encoding gene
 (glab) expressed in solid-state culture of Aspergillus oryzae";
 RL Gene 207:127-134 (1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=O-1013;
 RA Hata Y., Ishida H., Kojima Y., Ichikawa E., Kawato A., Suginami K.,

DE 01-OCT-2003 (TREMblrel. 25, Last annotation update)

OS Corticium rolfsii.

OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;

OC Aphyllophorales; Corticiaceae; Corticium.

OX NCBI_TaxID=39291;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AH09627;

RX MEDLINE=96158471; PubMed=8597548;

RA Nagasaka Y., Muraki N., Kimura A., Suto M., Yokota A., Tomita F.;

RT Cloning of Corticium rolfsii glucoamylase cDNA and its expression in

RT Saccharomyces cerevisiae.";

RL Appl. Microbiol. Biotechnol. 44:451-458(1995).

DR EMBL; D49448; BAA08436.1; -.

DR HSSP; P04064; ICGI.

DR GO; GO:0004439; F:glucan 1,4-alpha-glucosidase activity; IEA.

DR GO; GO:0016788; F:Hydrolase activity, acting on glycosyl bonds; IEA.

DR GO; GO:0005976; F:polysaccharide metabolism; IEA.

DR InterPro; IPR002044; CBD_4.

DR InterPro; IPR008291; Glu-a-gledc SBD.

DR InterPro; IPR000165; Glyco_hydro_15.

DR InterPro; IPR008928; Glyco_trans_6np.

DR Pfam; PF00686; CBM_20; 1.

DR Pfam; PF00723; Glyco_hydro_15; 1.

DR PRINTS; PR00736; GLHYDRLASE15.

DR ProDom; PD001568; CBD_4; 1.

DR PIRSF; PIRSF001031; Glu-a-gledc_SBD; 1.

DR Glycosidase; Hydrolase.

QO SEQUENCE 579 AA; 61548 MW; F556D83F9D1A9EC9 CRC64;

Query Match	50.3%	Score 1394	DB 3	Length 579
Best Local Similarity	51.2%	Pred. No. 8.8e-81		
Matches 294	Conservative 77	Mismatches 151	Indels 52	Gaps 11

```

QY 3 FRSLIATSLGVLCTGANTVANSKRATLIDMSLNEATAPARAIINNTGADGAWMSGADSGVIV 62
Db 2 FRSLIALA---ACAVASVSAOSASATYALITRESAAVAKGVICLNIISQCMSEGAASGVIV 58
QY 63 ASPSTNDPDEVYTWTRDSGLVLTLYVDLFRNG-DPSLSTLENIYSAOAIVOGISNPSGD 121
Db 59 ASPKTSBPDYLYWTRDSSLVFKMLIDPYTMGLDTTLTLIDEFASARATIIQOTSNSGT 118
QY 122 LSSGAGLGEPEKNVDETAYTGSWGRPORDGAPALRATAMIGGOMLIDNGYSTATDIWTP 181
Db 119 VSTG-GIGEPKFNIDETAFITGAKGRPGQDGAIRATITAMTYATLYNNGNITSYNTIMLP 177
QY 182 LVNRDLSVAQYMNQGTGYDIMEEYVNSFFITAVOHRALVEGSAFATAVG--SSCSWDS 239
Db 178 IIKLDLPVNSDMNQTTFTDLMEEVDSSFFTTAVOHRALVOCARATILIGQSSASTVSA 237
QY 240 QAPETILCYLQSFW-TGSFIFLANFPDSSRKGKQANTLISIRHPDEPAACDDSTFQPCSPRA 298
Db 238 TAPILCFLOGYMNWNTNGYTWANTGGGRSGKQANTILIAIHFPDASAGCSAATSQPCSDVA 297
QY 299 LANKEVVDSSRSIYTLINDGLSDSAVAVGVPEBPTYNNGNPMFLCTILAAEQLYDALYQ 358
Db 298 LANIKVYVDSFRSIYTTINSGISSTGAVTGRPEBSYVNGNPMFLCTILAAEQLYDALIV 357
QY 359 WDKQSGLEVTVDSLDFPKALYSDAATGYSSSSSTYSISIVAVKFPADGAPSIYETHAAS 418
Db 358 WKAAAGELNVTYSVLAFPOQFPSSITAGIYASSSSSYTILSDIOAFADFPDIYAKYTPS 417
QY 419 NGSMSQYDKSDGEQOLSARDLTWSYALLTLANNRRNSVVPASWGETSASVPGTCA--AT 476
Db 418 SGFLSEQYDKSTGAQDPSAANLTWSYAAALITVQANNGFTGASWG--AKGVSTSGSTGAT 474
QY 477 SAIGTYSSVYTT-----SWPSIVATG---GTTTAT-----PTGSGSVTST 514
Db 475 SPGSGSGSVEYTFPDVATTVYGVGNIYIGDVSELGNTFPANGVALSANYPTWSATIALP 534
QY 515 SKTT-----ATASKSTTTTRSG 531

```

Db	535	ADTTIGKYNNIDGSTVIMEDAI5NRIT7TPASG	568
RESULT 7			
Q9P4C5			
ID	Q9P4C5	PRELIMINARY;	PRT; 571 AA.
AC	Q9P4C5		
DT	01-OCT-2000	(TrEMBLrel. 15, Created)	
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DS	Glucosamylase.		
OS	Lentinula edodes (Shiitake mushroom) (Lentinus edodes).		
OC	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;		
CC	Agaricales; Tricholomataceae; Lentinula.		
OX	NCBI_TaxID=5353;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	Medline=20292826; Pubmed=10831434;		
RA	Zhao J., Chen Y.H., Kwan H.S.;		
RT	"Molecular Cloning, Characterization, and Differential Expression of a		
RL	Glucosamylase Gene from the Basidiomycetous Fungus Lentinula edodes.";		
FT	Appl. Environ. Microbiol. 66:2531-2535(2000).		
DR	EMBL; AF220541; AAF75523.1; -.		
DR	HSSP; P04064; ICA1.		
DR	GO; GO:0003824; F:catalytic activity; IEA.		
DR	GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.		
DR	GO; GO:0005976; P:polysaccharide metabolism; IEA.		
DR	InterPro; IPR002044; CBD 4.		
DR	InterPro; IPR008291; Glu-a-glc5d_SBD.		
DR	InterPro; IPR000165; Glyco_hydro_15.		
DR	InterPro; IPR008928; Glyco_transf_6hp.		
DR	Pfam; PF00686; CBM_20; 1.		
DR	Pfam; PF00723; Glyco_hydro_15; 1.		
DR	PRINTS; PR00736; GHYDRLASE15.		
DR	ProDom; PD001568; CBD 4; 1.		
DR	PIRSF; PIRSF001031; Glu-a-glc5d_SBD. 1.		
DR	SEQUENCE 571 AA; 61161 MW; A1BD1EC84BE97FF CRC64;		

Query Match Similarity	46.4%	Score 1286;	DB 3;	Length 571;
Best Local Similarity	47.9%	Pred. No. 6.7e-74;		
Matches 266;	Conservative 91;	Mismatches 150;	Indels 48;	Gaps 10;

[illegible]

```
QY 472 ---TCAATSAIGTY-----SSVTVTSWPSIVAGCTTTT-----ATPTGSGSVTSTS- 515
Db 492 NVYLTGAVDALBEDMDSTNDAILLSSANYPTWSVTVDLPGSDVQYKIKKDGSGTVWESD 551
QY 516 ---KTATATAKSTTT 527
Db 552 PNMEITTPANGTYAT 566

RESULT 8
Q8J0P8 PRELIMINARY; PRT; 609 AA.
ID Q8J0P8 AC Q8J0P8;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Glucoamylase Glam.
OS Mucor circinellioidea.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Mucor.
OX NCBI_TaxID=36080;
RN [1]
RE SEQUENCE FROM N.A.
RA Houghton-Larsen J., Pedersen P.A.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY168303; AN85206.1; -.
DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
DR GO; GO:0005976; F:polysaccharide metabolism; IEA.
DR InterPro; IPR005036; CBM_21.
DR InterPro; IPR00165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF03370; CBM 21; 1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PRINTS; PR00736; GLHYCOMLASE15.
DR PROSITE; PS00820; GLUCOAMYLASE; 1.
SQ SEQUENCE 609 AA; 64778 MW; 56DD54CD371CC7C1 CRC64;

Query Match 28.3%; Score 784; DB 3; Length 609;
Best Local Similarity 39.1%; Pred. No. 8e-42;
Matches 175; Conservative 73; Mismatches 158; Indels 42; Gaps 11;

QY 25 ATLDWSLNSNATVARTAILNNGADGAWSGAGSIVASPTDNDPYFTWTRDGLVL 84
Db 173 STSSWIDGQTSVSRVAMLRNINP-----AGAVTGFLAASLSTAGPDYFYAMTRDSALTS 227
QY 85 KTVLDLFRN---GDTSLSTIENYISAQAIYQGISNPSGDLSSGAGLGEPENVDERTAYT 141
Db 228 HVAAYDYNNTLLAGNSTLLGLKKNYVTFSLNSQTTSTVCN-----CLGEPKFNKSSGYS 281
QY 142 GSWGRPQRDGPALRATAMIGFGQWL--DNGYTSTATDIWPLVHNDSLRYVAQYNNQTYD 200
Db 282 GAWGRPQNDGPALRATAMIGFGQWL--DNGYTSTATDIWPLVHNDSLRYVAQYNNQTYD 200
QY 201 LMEEVNGSSFFTLAVQHRALVESAFAFVAGSS--CSWCDSQAPELICYLQSEF--TGSF 256
Db 342 LMEEVNGHVFYTLVWVRGVLKVGANFATRNDSSTRATYTTAASIKTKIDSFMNSNGQY 401
QY 257 ILAIFD-----SSRGKQANTL---LGSIHFPDEACDDSTPQCPSPRALANHKEVVD 308
Db 402 VVSQSVTGVSXKAGYDASVLIASNLGSL-----QDCFYTPGSDKMLATAVAESK 452
QY 309 FRSIYTLNDGLDSEEAVALVGRYPEDTY-----YNGNFWFLCTLAAPQLYDALYQMDKOG 363
Db 453 FASLYPINKNLPSYLGNSIGRYPEDTYNNGNSQGNPWFIAVGYAELYYRAIKETWNG 512
QY 364 SLEVTIVSLDFPKALYSDDATGT--YSSSSSTYSIYDAVKTFPADGFSIVETHAASNGSM 422
Db 513 GVTYTSISLNFPPKFKFSSAAVGTGYVGTSSFNLSLVQVAADAFFSTIKFHAATNGSM 572
QY 423 SEQYKSDGEBQLSARDLTWGAALLTAN 450
Db 573 SEQYGRADGLMTGARDLTWGAALLTAN 600
```

```
RESULT 9
Q727X9 PRELIMINARY; PRT; 579 AA.
ID Q727X9 AC Q727X9;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Glucoamylase (Fragment).
OS Rhizopus oryzae (Rhizopus delemar).
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Rhizopus.
OX NCBI_TaxID=64495;
RN [1]
RE SEQUENCE FROM N.A.
RA Chen G., Yu X.C., Jiang H.Z., Li M.G.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY359821; AA018643.1; -.
DR NON TER 1 1
FT CHAIN <1 26 glucoamylase.
FT CHAIN 27 85 glucoamylase.
FT CHAIN 86 411 glucoamylase.
SQ SEQUENCE 579 AA; 62151 MW; 75811D6936BB8751 CRC64;

Query Match 26.3%; Score 728; DB 3; Length 579;
Best Local Similarity 37.1%; Pred. No. 2.8e-38;
Matches 166; Conservative 70; Mismatches 170; Indels 42; Gaps 10;

QY 25 ATLDWSLNSNATVARTAILNNGADGAWSGAGSIVASPTDNDPYFTWTRDGLVL 84
Db 143 STSSWIDGQTSVSRVAMLRNINP-----PGSATGFLAASLSTAGPDYFYAMTRDSALTS 197
QY 85 KTVLDLFRN---NGDTSLSTIENYISAQAIYQGISNPSGDLSSGAGLGEPENVDERTAYT 141
Db 198 NVLYEYNTLLSGNKITLAVLKDYVTFVSQTSSTVCN-----CLGEPKFNKSSGYS 251
QY 142 GSWGRPQRDGPALRATAMIGFGQ--WLDNGYTSTATDIWPLVHNDSLRYVAQYNNQTYD 200
Db 252 GAWGRPQNDGPALRATAMIGFGQ--WLDNGYTSTATDIWPLVHNDSLRYVAQYNNQTYD 200
QY 201 LMEEVNGSSFFTLAVQHRALVESAFAFVAGSS--CSWCDSQAPELICYLQSEF--TGSF 258
Db 312 LMEEVNGHVFYTLVWVRGVLKVGANFATRNDSSTRATYTTAASIKTKIDSFMNSNGQY 401
QY 259 ANFDS-----SSRGKQANTL---LGSIHFPDEACDDSTPQCPSPRALANHKEVVD 308
Db 372 IQVSQSVTGVSXKAGYDASVLIASNLGSL-----DDGFTTPGSEKILATAVAESD 422
QY 309 FRSIYTLNDGLDSEEAVALVGRYPEDTY-----YNGNFWFLCTLAAPQLYDALYQMDKOG 363
Db 423 FASLYPINKNLPSYLGNSIGRYPEDTYNNGNSQGNPWFIAVGYAELYYRAIKETWNG 512
QY 364 SLEVTIVSLDFPKALYSDDATGT--YSSSSSTYSIYDAVKTFPADGFSIVETHAASNGSM 422
Db 483 GVTYTSISLNFPPKFKFSSAAVGTGYVGTSSFNLSLVQVAADAFFSTIKFHAATNGSM 542
QY 423 SEQYKSDGEBQLSARDLTWGAALLTAN 450
Db 543 ABEFDRITTGASTGARDLTWGAALLTAN 570

RESULT 10
Q8TFE5 PRELIMINARY; PRT; 515 AA.
ID Q8TFE5 AC Q8TFE5;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Glucoamylase precursor (EC 3.2.1.3).
GN GUU 0111.
OS Saccharomyces fibuligera (yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
```

OC Saccharomycetales; Saccharomycopsidaceae; Saccharomycopsis.
 OX NCBI_TaxID=4944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IFO 0111;
 RA Hostinova E., Solovicova A., Janacek S., Gasperik J.;
 RT "Raw starch degrading glucanase from Saccharomycopsis fibuligera:
 molecular cloning and expression in yeast."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ311587; CAC8369.1; -;
 DR GO: GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
 DR GO: GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
 DR GO: GO:0005976; P:polysaccharide metabolism; IEA.
 DR InterPro: IPR000165; Glyco_hydro_15.
 DR InterPro: IPR008928; Glyco_trans_6hp.
 DR Pfam: PF00723; Glyco_hydro_15; 1.
 DR PRINTS: PRO0736; GLUCRASE1.
 DR PROSITE: PS00820; GLUCOMYLASE; 1.
 DR Signal; Hydrolase; Glycosidase.
 FT SIGNAL 1 26
 FT CHAIN 27 515 GLUCOMYLASE.
 SQ SEQUENCE 515 AA; 57423 MW; A2F27D9E8A42AE0 CRC64;
 Query Match 24.3%; Score 674; DB 3; Length 515;
 Best Local Similarity 33.7%; Pred. No. 6,7e-35;
 Matches 163; Conservative 82; Mismatches 177; Indels 62; Gaps 16;
 QY 24 RATLDWLSNERTATATLANNIGADGAWSGADSGIVASPESTDNDPYTTTROSGLV 83
 DB 41 RSNFTQWHEQPAVSWYLLQNDYPEGQFSAKPGVVAFPSTSEDPDYQWTRDTAT 100
 QY 84 LKTLVLDLFRN---GDTSLSTIENYISAQIVOGISNPSGDLSS--GAGLGEKFNVD 138
 DB 101 FLSLIEVEDHSSNTTLAKVVEYISNTYTLORVSNPSGPNHSDGEGEKFNVDDT 160
 QY 139 AYTSGWRPQRDPALRATAMIGF-----GQWLL--DNGYT--STATDVMPLVRND 186
 DB 161 AYTASWGRPDNDPALRAVAISRILYNAVAKHNNKLLAGQNGIPYSSADIDYWKIKPD 220
 QY 187 LSVVAQYNNOTGVDLMEENVSSFFITAVOHRALVEGSAFATAVGSS--CSWCDSPAET 244
 DB 221 LQVVSHTWSTSGDLMEENOGTHFTALVOLKALSYGIPLSKTYNDPGFTSWLEKQDAL 280
 QY 245 LCVLOSFWTGSFILANFD-----SSRSGKDANTLLGSIHTEPDEACDSDTOPCSP 296
 DB 281 NSYINS---SGFVNSGKHAIVESPOLSSRGGLDSATYIALITHD--IGDDDTYTPFN-- 333
 QY 297 RALANHEVVDSP-----RSIYTLNDGLSDSEAVAVGRYPEDTY-----YNGNPFPLC 344
 DB 334 ---VDNSVYLNSLYLLVDNKNRYKINGNY--KAGAAVGRYPEDYVNGVGTSGGNPQOLA 388
 QY 345 TLAAEQILDALYQWOK--QGSLEVTDVSLDFKALYSDA--TGYTSS-----SSST 393
 DB 389 TAYAGOTFTYLANSLKNNKGLVIEKLANVDLYNSFTADLSKIDSSYASKSLTLYGSDN 448
 QY 394 YSSIVAVTFAPGAFVSIVETHAASNGSMEQDSDGEQLSARDLTWSYAALTANNR 453
 DB 449 YKNVLSLQDFGSLKVLKVLHDIDNGQLTEINRTYTFQAGAVSLTWSGSLSANRAR 508
 QY 454 NSVY 457
 DB 509 NKLI 512
 RESULT 11
 Q96210 PRELIMINARY; PRT; 599 AA.
 AC Q96210;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein ST2017.

OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OX Sulfolobus.
 RN NCBI_TaxID=111955;
 RP SEQUENCE FROM N.A.
 RC STRAIN=JCM 10545 / 7;
 RX MEDLINE=21456156; PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudo H., Yamazaki J., Kushiida N., Oguchi A.,
 RA Aoki K.-I., Maeda S., Yanagii M., Nishimura M., Yamagishi A.,
 RA Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermophilic
 Crenarchaeon, Sulfolobus tokodaii strain7."
 RL DNA Res. 8:123-140(2001).
 DR EMBL: AP000988; BAB67116.1; -;
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
 DR GO: GO:0005215; F:transporter activity; IEA.
 DR GO: GO:0005976; P:polysaccharide metabolism; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR000165; Glyco_hydro_15.
 DR InterPro: IPR008928; Glyco_trans_6hp.
 DR Pfam: PF00723; Glyco_hydro_15; 1.
 DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
 DR Hypothetical protein; Complete proteome.
 SQ SEQUENCE 599 AA; 69661 MW; F8B94CAE731D119 CRC64;
 Query Match 6.7%; Score 187; DB 17; Length 599;
 Best Local Similarity 22.3%; Pred. No. 0.001;
 Matches 102; Conservative 58; Mismatches 151; Indels 146; Gaps 23;
 QY 29 SWL-----SNEATVARTAL-----NNIGADGAWSGADSGIVASPESTDNDPYFTWTR 78
 DB 235 AMLVAKARDVLSVRRLILIAHWQNN-----GALPALDITDINRKNKT-----YNYVMHR 286
 QY 79 DSGLVKTLVLDLFRNGDTSLSTIENYISAQIVOGISNPSGDLSSGAGLGEKFNVD 138
 DB 287 DAAF-----ASIALTLVGY---QPIRNLFPNFKPLINGFLEQ--KYTCD-- 327
 QY 139 AYTSGWRP-----QRDPALRATAMIGFQWLLDNGYTTSTANDIYVPLVRNDL 187
 DB 328 ---GNMGSTWHPNPRSPIQED-----BTALMLYALVWHSFRFTD--IDFAPPLAPVY 377
 QY 188 SYVAQY-----NOTG-----YDLMEENVSSFFITAVOHRALVEGSAFATAVGSSCSWC 237
 DB 378 KKAIEPLVSRDEETLPLPSIDLMEERLGTHTFTSLAVYAGLSAYKRAFEFGDE--NLK 436
 QY 238 D--SQAPILCYLOSFWTGS--FILANFDSRSGK--DANTLLGSIHTEPDEAACDSTF 291
 DB 437 DKYLTAAENVKKGLERFYVGDHFAITYEDNSIDKTVDASTLPAASIL-----GPF 486
 QY 292 QPCSPALANHEVVDSPFSITLNDGLSDSEAVAVGRYPEDTY-----NGNPFPLCTLA 347
 DB 487 DPKDPRVVISNRKVEKEK-----NINNG-----TARVENMYLQODEKSNAMFTTTLW 534
 QY 348 AAEQLYDALYQWOKQSLSEVTVSLDFKALYSDAATGYSSSSSTYSSTVDAVKTADG 407
 DB 535 LAQY-----ILBNKKEKAKK 550
 RESULT 12
 Q9HL88 PRELIMINARY; PRT; 636 AA.
 AC Q9HL88;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein ST2017.

DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein Ta0342.
GN Ta0342.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatia; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
Mewes H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
acidophilum."
RL Nature 407:508-513(2000).
DR EMBL; AL445064; CAC11486.1; -
DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
DR GO; GO:0005976; F:polysaccharide metabolism; IEA.
DR InterPro; IPR000165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00723; Glyco_hydro_15; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 636 AA; 73269 MW; FFA4656CCDE9AFD CRC64;

Query Match 6.7%; Score 185.5; DB 17; Length 636;
Best Local Similarity 19.0%; Pred. No. 0.0014;

Matches 86; Conservative 56; Mismatches 119; Indels 171; Gaps 15;

QY 57 DSGIVASPSPTD-----NPDPYFTWTRDSGLVLTIVDLFRNGDPT--LLSTIENYISAQ 109
DB 281 DLGAIASCSODILKSHDGYVYWPADASMAAYAL-SIGHSPTARFALMDSISEE 339
QY 110 AIYOGISNPSGDSGSGGLGEPKKNVDETATGSGWRPQDGPL-----RATMIGFGQ 164
DB 340 GYLYHKYNDGKIAS-----SMLPVMNGKSIYPIQEDETALVYVVAL 381
QY 165 WLDNGYSTATDVIWPLVRNDLSVYAQYNN-----QTCYDLWEE 204
DB 382 WEYFRKY-----NDIGFTAPYERLITRAADFMWTVNDNGPKPSFDLWEE 428
QY 205 VNGSSPFTIAVQHRALVEGSAFATAVGSSCWCDSPAELCYLQSFMTGSFILANPDS 264
DB 429 RYGHIAVTATVVAALKAASNFANVFG-----DPD 458
QY 265 RSGKDANTLLGSIHTPPEACADSTFQCSPRALANKE--VDSFRSITYTLNDGLSDS 322
DB 459 LSEKYENAAERMYAHAFDERFYSEDTGY--ARAIIDGKDPFTVDSALTSILV--FGMKDA 514
QY 323 E-----AVAVGRYPEDTYN-----GNPFLCTLAAAEQLYDA 355
DB 515 DDEKVIISTMGRISDLWVNGVGIAKYQNDRIKRVKDDPSVPGNPMITILLMAH-RYYMR 573
QY 356 LYQWMDKQSLVETDVSLLDFKALYSDAATGTYSSSSSTYSIYDAVATFADGFVSIYETH 415
DB 574 FGPFEKAMNL-----IQWKSH 590

RESULT 13

ID Q8XDG4 PRELIMINARY; PRT; 1588 AA.
AC Q8XDG4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Putative adhesin (Hypothetical protein).
GN Z5029 OR ECS4480 OR H161.

OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Rafaeli G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
R. Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
R. Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
R. Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
R. Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
R. Kohara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 0509952;
RA Makino K., Yutsudo C.H., Yokoyama K., Kubota Y., Kimura S.,
R. Shingawa H.;
RT "O157 specific gene similar to H. influenzae adhesin gene.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB005587; BAG58749.1; -
DR EMBL; AB002566; BAB37803.1; -
DR EMBL; AB036416; BAB87814.1; -
DR PIR; A86036; A86036.
DR PIR; H91188; H91188.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM_
DR InterPro; IPR005594; Yada,
Pfam; PF05658; Hep_Hag; 13.
DR Pfam; PF05662; HIM; 12.
DR Pfam; PF03895; Yada; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 1588 AA; 160150 MW; B2BA4E06FF28DEC CRC64;

Query Match 6.5%; Score 181; DB 16; Length 1588;
Best Local Similarity 21.7%; Pred. No. 0.0092;

Matches 136; Conservative 95; Mismatches 217; Indels 178; Gaps 30;

QY 15 TGLANYSIKRATLDSWLSNATATVARTALLNNIGADG-AWMSGADSGIVASPSPTNDPDI 73
DB 576 TNLANTSNIAITNTWISN-----LIEVTNLGEDALKM--DYONGFTNAGHETTSKI 628
QY 74 YTWTRD-----SGVLKTLVLFPMGDTSLSTIENYISAQAIYOGISNPSGDL 122
DB 629 -TNVKGDLITGSGTDAVNGSLKTTNDVATNTNTNIAITNTN-----ISNLIEIV 677
QY 123 SSGAGLGEF--KENVDETATGSGWRPQRDGPALRATAMIGFGQMLDNGYSTADIW 180
DB 678 TN--LGEDALKWDKNDNGVFTAHG--NNTASKITN-----ILDGIVTATSSDAIN 723
QY 181 PLVRNDL-STVAQYW-----NOTG-----YDLMEVNGSSFTIAVQHRALVEGSAFA 227
DB 724 GSQLYDLSSNIATYFGGNASVNTDGVFTGPTYIGE--TNYNVGDALAAI--NSSFS 777
QY 228 TAVSSCSWCDSPAELCYLQSFMTGSFI--LANPDSRSRQDA--NTLLGSIHTF--- 280
DB 778 TSLGDALW-DATAGKFSAGHTNGDASVITVDADGEISDSSDANVNGSLDHGVSSVVD 836
QY 281 ---DPEACDSTFQCSPRALANKEVVDSFRSI-YTLNDGLDSEAVAVGRYPEDTY 335


```

Db      837 ALGCGAEVNAVADGTTTAPTYTIANADYDNDVADALNADITLDDAL-----LMDADAG 887
QY      336 YNCPMPFLCTLAABEQLYDLYQMDKQGLSEVTVDSLDFPKALYSDAATG----- 385
Db      888 ENG-----AFSAHKGDKRTASV-ITNVANGAISAAASDAINGSOLYTTNKYI 933
QY      386 -----TSSSSSTYSISIVDAVKTF-----A 405
Db      934 ADALGGDAEYNADGTTTAPTYTIANAEYNNVGDALDLDLNDALMDETANGAGAYNASH 993
QY      406 DGFVSIETHAASNGSMSE-QYDKSDGEQLSA-----RDLTWSYAL 446
Db      994 DGVASIIITN--VANGSISEBSTDVANGSOLNATNMMEQNTQIINQAGNTDATYIQENG 1051
QY      447 LTNRRNRNSVVPASMETASSVPGTCATSAICTYSSVTVWSPSIVATGCTTTATPT 506
Db      1052 AGINIVRTDGDGLAFPDASQGVGAT----AIG-YNSVAKGDSVAIGGGSYSDVTGI 1105
QY      507 GSGSVTSTKTKTATASKSTTTTSGM 532
Db      1106 ALGSSSVSRVIAKGSRDISTENG 1131

RESULT 14
Q9C105 PRELIMINARY; PRT; 1236 AA.
ID 09C105;
AC 09C105;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Putative glucosylase I (Alpha-1,4-glucan glucosidase), extracellular
  starch-degrading enzyme, by similarity to S. cerevisiae STA1, contains
  chitinase family signature.
DE SPAP1E7.04C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Wood V., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL/AJ590605; CAC36921.1; -
DR GeneDB_Spombe; SPAP1E7.04C; -
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001917; AminoTrans-II.
DR InterPro; IPR001223; Glyco_hydro_18.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
SQ SEQUENCE 1236 AA; 123387 MW; 5A2D3JA30B87CDD8 CRC64;

Query Match 6.3%; Score 175; DB 3; Length 1236;
Best Local Similarity 21.5%; Pred. No. 0.016;
Matches 189; Conservative 91; Mismatches 242; Indels 138; Gaps 21;

```

```

Db      693 SSSSSPSPSSSVLT-----YASLSNI-----THSLSLTPM-----SSSSAI 732
QY      202 WEVNSSSSEFTIAVQRAALVEGSAFATVAGSS--CWCDSQAPELICYSFTGSEFIIA 259
Db      733 PTVNSSTLIT-ASSNTLLSITSSSAIVSSTVSNISNLPASATASSQSOLNSSTLA 791
QY      260 N--FDSSRGKQKANTLLGSIHTFPDPAACDDSTFOCPSPALANHEVDSFRIYTLND 317
Db      792 TSLYLSSSSSRTISTSTNEYNTSFLAPVSSSTSSSTTSLAANKV--NSNISTSLN- 848
QY      318 GLSDSAVAVGRPEEDTYNGNPMFLCTLAABEQLYDLYQMDKQGLSEVTVDSLDFPKA 377
Db      849 -LESTSVSTAVTDSVTS-----TTLATSQ-----GPSVSSSSLSSTTS 890
QY      378 L-----YSPAAGTYSSSSTTSYVD-----AVKTFADPFYSLVER 414
Db      891 LSTSIPTSSVAPAVTSTSETSSVVGSGTDSATSSMTAETSSSAITSSVAASVPTSS 950
QY      415 HAASNGSMSEQYDKSDGEQLSARDLTWSYAALLTNRRNSVVPASMGETSASSVPG--- 471
Db      951 SSASSSSSEVDPTSAATGSS-TSSITATASVSGSSSTSSVATATSDSTSSIAASV 1009
QY      472 TCAATSAICTYSSVTVWSPSIVATGCTTTATPTSGSVT-STSKTTATASKSTTTTRS 530
Db      1010 TGSSTSVAT-ASVTDSTSSVATASATDSTSSIAVASVTSSTSSVATASATDSTSS 1068

RESULT 15
Q97BM7 PRELIMINARY; PRT; 659 AA.
ID Q97BM7;
AC Q97BM7;
DT 01-OCT-2001 (TEMBLrel. 18, Created)
DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Glucosylase.
DE TV0428 OR TVG0415361.
GN Thermoplasma volcanium.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
  Kawashima T., Yamamoto Y., Watanabe K., Yamazaki M., Kanehori K.,
  Nunoshida T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
  "Archaeal adaptation to higher temperatures revealed by genomic
  RT sequence of Thermoplasma volcanium.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL; AF000992; BAB59570.1; -
DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
DR GO; GO:0005976; P:polysaccharide metabolism; IEA.
DR InterPro; IPR00165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00723; Glyco_hydro_15; 1.
KM Complete proteome.
SQ SEQUENCE 659 AA; 76598 MW; 36101F5405F13B33 CRC64;

Query Match 6.3%; Score 174.5; DB 17; Length 659;
Best Local Similarity 19.7%; Pred. No. 0.0071;
Matches 90; Conservative 49; Mismatches 158; Indels 159; Gaps 16;

```

[illegible]

	RESULT 16					
ID	Q8J3Z2	PRELIMINARY;	PRT;	990 AA.		
AC	Q8J3Z2;					
Dt	01-JUN-2003	(TREMBLrel . 24,	Created)			
Dt	01-JUN-2003	(TREMBLrel . 25,	Last sequence update)			
Dt	01-OCT-2003	(TREMBLrel . 25,	Last annotation update)			
DE	Orf, hypothetical protein. SF3641.					
OS	Shigella flexneri.					
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.					
OX	NCRF_Taxid=623;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=301 / Serotype 2a;					
RX	MEDLINE=22727406; PubMed=12384590;					
RA	Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang T., Liu H., Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J., Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S., Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y., Yu J.;					
RA	"Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157.";					
RL	Nucleic Acids Res. 30:4432-4441(2002).					
DR	EMBL; AE015372; AAAA5088.1; - InterPro; IPR005594; YadaA. Pfam; PF03895; YadaA; 1.					
KW	Hypothetical protein; Complete proteome. SEQUENCE 990 AA; 100459 MW; AADBDS2F9940E92B CRC64;					
SQ						
	Query Match	6.3%; Score 174.5; DB 16; Length 990;				
	Best Local Similarity	21.3%; Pred. NO. 0.012;				
	Matches 128; Conservative	93; Mismatches 201; Indels 179; Gaps 28;				
Cy	43 LNMGADG-AAMVSGADSGIVVASPETDNDPIFYTRD-----SGLVKTLVDL 90	:	:	:	:	:
Dd	1 MTNLGEDALKM--DKDNQGVFTAAHTEFTYSKI-TNVKDGDLITGSTDAVNSSQLKTINDA 57	:	:	:	:	:
Cy	91 FRNGDTSLSTEINENYISAQAIVQGINSFGDSLGSAGLEGP--KNVDETAYTGSWGRRQ 148	:	:	:	:	:
Dd	58 VAVNTTNIAIATINTN-----ISNLTETVTN---LGSDALKKWDKNVGFTAHG-- 101	:	:	:	:	:
Cy	149 RDGPALTATAMIGFCQMILIDNGYSTATDIWPLVRNDL-SYAQYM-----NQTG--- 198	:	:	:	:	:
Dd	102 --NNHTASKITN-----ILDGTIVMTTSDDAINSQLYLSSLAIATYYPGGNASVANVDGYPT 153	:	:	:	:	:
Cy	199 ---YLMEEVNGSSFETTIAVOGHRALVEGAFAPATVAAGSSCWCDSCAPEILCYLOSFWTGS 255	:	:	:	:	:

```

Db      154 GPYYKIG-----TNYNNGDALAAI--NSSFETSLGDLMLW-DATAGKRSAAKHGTMGDAS 206
Qy      256 FI--LANPSSRRSGKDA--NTLLSGIHTF-----DPEACDDSTFQCCSPRALANHKE 304
Db      207 VITVDADDERIDS$SDAVNGSOLHGVSSVVDVALDGGCAEVNADGTTIAPTYYTIANADYDN 266
Qy      305 VVD$RSLTYTLNDG--LSDSEAVNAGRYPEDTYNNGNPNWFLCTLAAEQLYDALYQWMDQ 362
Db      267 VGDALNALIDTTPDDALLWMDADGENG-----AFSAHAGKDKT 303
Qy      363 GSLEVTVDLSLDFKALYSDAATG-----TYSSSSST 393
Db      304 ASV-ITTVANGAISA$SDALINGSQLYTTNKYIADLGGDAEVNADGTTIAPTYYTIANAE 362
Qy      394 YS$IVDAKTF-----ADGFVSIVETHAASNGSMSE-QYIKSDG 431
Db      363 YNNVGDALDALDNLALMDKTANNGAGAYNASHDGA$IIITN--VANGSISEB$TAVVNG 420
Qy      432 EOLSA-----RDLTWSYALLTLANNRNSVSPASWETSASVPGTC 473
Db      421 SOLNNTNMIEONTIOIINOLAQNTATYIIEENGAGINIVYRTDNDLA$FNDAS$SGVGA 480
Qy      474 AATSAI$TYSSSVTVLWSP$IVATGTTIYATP$G--SGSVTS$KTTATASKT$ITTR$G 531
Db      481 VGVNA-----VASGASSVAIGNSS$TVDTGIALGSSSVSSRVIAK$GRD$T$VTEMG 532
Qy      532 M 532
Db      533 V 533

```

ID	Q8TGE1	PRELIMINARY;	PRT;	1713 AA.
AC	Q8TGE1;			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)			
DE	Cell wall protein Awalp.			
GN	AWAL			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21914116; PubMed=11916725;			
RA	Shinoi H., Sakamoto K., Okuda M., Atchi R., Iwashita K., Ito K.;			
RT	"The AWAL gene is required for the foam-forming phenotype and cell			
RT	surface hydrophobicity of sake yeast."			
RL	Appl. Environ. Microbiol. 68:2018-2025(2002).			
DR	EMBL; AB071164; BAB85832.1; -			
DR	GO; GO:0006950; P:response to stress; IRA.			
DR	InterPro; IPR000992; SRP1_T1P1.			
DR	Pfam; PF00660; SRP1_T1P1.			
DR	PROSITE; PS00724; SRP1_T1P1; 1.			
SQ	SEQUENCE 1713 AA; 16683 MW; 9321A296B5B86404 CRC64;			
Query Match 6.3%; Score 173.5; DB 3; Length 1713;				
Best Local Similarity 22.6%; Pred. No. 0.031;				
Matches 125; Conservative 91; Mismatches 281; Indels 55; Gaps 15;				
QY	7 LALSLVLTGLANVYSKATLDSWLSN-EATVVARILAININIGADGAWYSGADSLIVASP 65			
DB	393 LAVSSTVATSTDSVSPNIPSEIISSESSSTALTSGSSASATSGSSVSGSTASATSGSS 452			
QY	66 STDNDVRYETWRDSDGLVLTIVDLFRNGDPSLL-----STLENYISAQATVQGISNPSG 120			
DB	453 ASGSS-----ATGEGSSVSGSTGATGEGSASSVSSGSGVTGSSSSSSASBSSTLQ-SG 506			
QY	121 DLSSGAGLGEPEKNVDF--TAYTSGWGR-----PQDGPALRATAMIGFGQMLDN 169			
DB	507 -TAGSSSVTSSTGVTQGGSSVSGSSASAPGISSTIPQSTSSASTAGSITSGTISRT 565			

```

QY 170 GYTSFATLTWPLVANDLSYVAQYNNQGTGYDLMEENVGSSFFLIAYOHRLVGSAPATA 229
Db 566 SGGSSATE-----SGSSVSGSSSATEGSSVSGS--TSATEGSSVSGSTSTATE 612
QY 230 VGSGCCSWCDSCAPEILCYLOSFWTGSFIIPANFDSNRGKDANTLLGSIHTFDPEAACDS 289
Db 613 SGGSSASGSSSATES-----GSSVSGSTATEGSSVSGSTSTATEGSSSAGSSSATESGS 667
QY 290 TFC-PCSPRALANHEVVDSPRSITYTLNDGLDSEAVANGRYEDITYGNPFELCTLA 348
Db 668 ASSVPSGSSGSSVTEGSSSSSASBSSTITQSGTASGSSASSTSGSVTQSGSSVSGSSAS 724
QY 349 AEQLDALYOMDKQGLEVTVDVLDLPEFKALYSPAANGTSSSSSTYSIIDAVKTFADGF 408
Db 725 APGISSTIPQSTSSASTASGSTITGTLTITSGSSSATESGSSASGSSSATEGSSVSGS 784
QY 409 VSIIVETHAASNGSMSEYOYDKSDGEOLASDLTWVYALLTANNRRNSVVPASWGET---S 465
Db 785 TSATESGSSVSGSTSTATEGSSASGSSSATESGSSVSGSTSTATEGSSSAGSSSATESGS 844
QY 466 ASSVGTCAATSIGTYSVYVYTSMPDSIYATGGTTTAAPTPGGGSYTSISKTTAATASKTS 525
Db 845 ASSVPSGSSGSSVTEGSSSSSS---ASBSITQGTASGSSASSTSGSVTQSGSSVSGSSAS 901
QY 526 TT---TRSGMSL 534
Db 902 TSGSVTQSGSSV 913

```

RESULT 18	
Q97ZD0	
ID Q97ZD0	PRELIMINARY;
	PRT; 622 AA

DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Glucan 1,4 alpha glucosidase (Glucocamylase) (EC 3.2.1.3).
GN SS60990.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_Taxid=2287;
[1]
RN sequence FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332286; PubMed=11427726;
RA Sme Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awaeyer M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moers A., Etienne G., Fletcher C., Gordon P.M.K.,
RA Helkamp-de Jong I., Jeffries A.C., Kozeira C.J., Medina N., Peng X.,
RA La-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charalabos R.L., Doolittle W.F., Duguet C.W., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.,
RT "the complete genome of the crenarchaeon Sulfolobus solfataricus P2.".
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AB006718; AAK4162.1; -.
DR PIR; G90250; G90250.
DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005976; P:polysaccharide metabolism; IEA.
DR InterPro; IPR000165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_trans_6tp.
DR Pfam; PF007923; Glyco_hydro_15; 1.
KW HYDROLASE; GLYCOSIDASE; Complete proteome.
KW SEQUENCE 622 AA; 72052 MW; C7B81FD21AB8F CRC64;

	Query Match	6.2%	Score 172;	DB 17;	Length 622;
	Match Local Similarity	21.7%;	Pred. No. 0.0095;		
	Matches	97;	Conservative	165;	Indels 130; Gaps 19.
Cy	37 VARTALNIGADGAWSGADSCGIIVVASBSTNDPPIFYTWTRDSGLVLTIVYDLFRNGDT	96	:	:	:

```

Db 268 VSLVFINNHMDVNSIIASSPSFSFV----K1YGD9YQYCWPRDAIAAVAL-DLAgYKEL 322

QY 97 SL--LSTIENYISAQAVIQGISNPSGDISSGAGLGEKPFVNDETAYTGSWGRF-QRQCPA 153
Db 323 ALKHFOFISNANSEGFLYKHKNPPTLASS---WHPVY-----YKGRIYVIOEDETA 373

QY 154 LRATAMTGFQWLLDNGYTSATDT-----VWLVVRNDLSYVAQYNNQGYDLMW 203
Db 374 LEV-----MAIASHY-EKYEIDIDELPLYKKFVKPALKFPMFMEGELKRPSPDLWE 424

QY 204 EVNASSPFTTAVQHRALVEGSAFATVAVSSSCWCD5QAPELICYLOSFWTGSFILANPFS 263
Db 425 ERYGIIHYITVSTVYGALTKGAKLAYDGE-----ILSEDS 461

QY 264 SRGKDNANTLLGS-----IHTDPPE---AACODSTFQPCSPRAL-ANHKEVNDSPFS 311
Db 462 DTSGILKGMVLRKMTYNGRPFVRRIDEENNODLTVSSLSYAPFFGLVANDNDXIMTINE 521

QY 312 I---YTLNDGLSDBEAVAVGRYPRDYTY---NGNPWFCLTAAAEQLDALYQMPKQGS 364
Db 522 IESRLTVNGGII-----RYENDMYORKKQPPMPWITTL-----W----- 566

QY 365 LEVTDVSLDFPKALYSDAATGYSSSSSTYSISYAVATFADGFSIVETHAASNGSMSE 424
Db 557 -----LSIYYATIND-KKAKAEYIKWVNRRLPTGFLPE 589

QY 425 QYDKSDGEQLSARDLTWMSYALLTFANNR 452
Db 590 QVDETEPEPFSVTPLVWMSHAEPITAIANK 617

```

RESULT 19
Q88XB6
ID Q88XB6 PRELIMINARY; PRT; 3360 AA

DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Cell surface SD repeat protein precursor.
GN SDR OR LP 1303.1 OR LP 1303A.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RX [1]
RN SEQUENCE FROM N.A.:
RP STRAIN=NCIMB 8826 / WCFS1;
RC MEDLINE=22480296; PubMed=12566566;
RX Kleeberdezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.",
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995 (2003).
DR EMBL; AL935255; CAD63798.1; -
DR InterPro; IPR003343; Big_2.
DR Pfam; PF00368; Big_2; 1.
KW Complete proteome.
SQ SEQUENCE 3360 AA; 332641 MW; BBF2411DF439F446 CRC64;

```

Query Match      6.2%; Score 171; DB 16; Length 3360;
Best Local Similarity 20.9%; Pred. No. 0.11;
Matches 136; Conservative 75; Mismatches 228; Indels 212; Gaps 29;

QY      9  LSGVLCGLAN--VISKRAATDSWLSNE--ATVARTAIINNIGADGAW-VSGADSGIYVA 63
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      567  LAAFSTFDVANEQVSSNAILTTSTAQSVSTGCTVITTKNVGDSGVSQGTFL--YVM 624

QY      64  SPSTNDNDPYFTYTRDGLVLKTVLDLFRNGDTSILSTENIYISAQAIVOG----- 114
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      625  YVSTVDGR---TWTTISGATSATY-----TTATTSIGTY-YQAVNSKRTAAILTL 671

```

```

QY 115 -----ISNPSGD-----LSSGAGLGEPKENV 136
DB 672 SSFRAAGNVITVTNPEGYTEIASSVAQANSASVATSQAFAASASLASSANLT 731
QY 137 ---ETAYT-----GSGWRPQDGPALRATAMIGFQWMLDNGTSTATDITWPLV 183
DB 732 GSLSAYTTSQANSYNMASSASVASEASATATMASTASASYEAGATSN---IACM 788
QY 184 RNDLSYVAQWNTGYDLMEEVNCSFFITAVQHRALVEG-----SAFATVAGSSC 234
DB 789 ASSYASVATEYASASASAEAS--LDTALKNVAVDEGDDTADYLAASVAMAASA 846
QY 235 SMCDSQAPELICYQSWTGSFIILANPDSR--SGKANLTGSIHITPDPAACDSTFQ 292
DB 847 NVAGDQASLASV-----AANANAYNVAGSFFNTATSSV-----NAAASASAVA 892
QY 293 PCSPRALANHEVVDSPRSITTLNDGLDSEAVAV--GRYPEDTYNGNPMFLCTLAFAQ 351
DB 893 ASSYAAA-----SEASLASIYADGYNV-----AALQO 923
QY 352 LYDAL---YQWDK-----QGSLEVTDVSLDFKALYDAPATGVSSSSSTYSIVDAVK- 402
DB 924 SYALRLTYATDQANMAASASAITQTVQASGYNSIATASSAASAEASAGNALSIV 983
QY 403 -----TPADGFVSIVETH-----AASNGSMSEQYKSDGEQLSARDLT 440
DB 984 ASSISAKYASNTLTGSSYVQNTNNYAVTTLGAAAAATSAEMAQYTSQASASVAAADYA 1043
QY 441 W-----SYALLTANNRBNVSPASNGETSASV-----PGT 472
DB 1044 MASSYTNMAASAEASISYALSTSN-----AALGSTASNAVNTYNNNTNSYKSTNT 1095
QY 473 CAATSAIGTVSSVTVTSMPSIV--ATGCTT--TATPTGSGSVTSKTT 518
DB 1096 GNATTILIGIVSGMGTQPAVAVTSGGFTTLASSALGLGVALSCKT 1146

RESULT 20
ID 076602 PRELIMINARY; PRT; 1275 AA.
AC 076602;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Hypothetical protein.
GN H02F09.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showman R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL Mature 368:32-38 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Geisel C., Harmon G.;
RT "The sequence of C. elegans cosmid H02F09.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB077538; AAC64622.1; -.
DR F01; T33369; T33369.
DR WormPep; H02F09.3; CE19479.
KW Hypothetical protein.
SQ SEQUENCE 1275 AA; 122924 MW; CB59B70C05959E25 CRC64;

Query Match
Best Local Similarity 22.4%; Pred. No. 0.032;
Matches 129; Conservative 87; Mismatches 276; Indels 85; Gaps 19;

QY 1 MSFRLALSGVCTGLANVSKRATLDSWLSNEATVA--RTALINIGADGAVGADSG 59
DB 673 VSGSSVSTGSGSTSTAGASVSGSTGVSDBSSTISDSTGTMAGSTESTVGSVS 732
QY 60 IVASPESTNDPYFYTWTRDS---GLVLTLDV-LFRNGDTSLSTIENYISAQAVQGI 115
DB 733 TVGSGTSGTSPSTMSASTGTNTPGSTESTITDGSVSGSTGSTGNTNPGSTDSSTGI 792
QY 116 SNPSG-DLS--SGAGLGEPFNDERTAYTSGMGRPQDGPALRATAMIGFQWMLDNGYT 172
DB 793 STVSGSLSTISGSGTSGVSGSSDMVTSTGSSPGSTESTVSGASTMS-----P 842
QY 173 STATDIWPLVRNDLSVAQYMNQ--TGYDLMEEVNCSFFITAVQH-----R 218
DB 843 STGSSVSTSGSSVSTVSGSTSSSTTQ---STVSSSVSTVSSBETISQSTTTGE 899
QY 219 ALVEGSAFATVAGSSGSCWCSQAPELICYQSWTGSFIILANPDSRSGKANLTGSIH 278
DB 900 STVFGSGTATGSGSTMSASTGTDPGSTESTITGSGTVG--ESTVSGSTGSTITEG-- 955
QY 279 TPPEAACDSTFQ---CSPRALANHEVVDSPRSITTLNDGLDSEAVAVGRYPEDTY 335
DB 956 -----STISSTMTVGVSTGSGTITGSGTSGSTRVSGSTVSGSTESTVSGSTEST- 1009
QY 336 YNGNPMFLCTLAFAEQLYDALYQMDKQSLLEVTVSLDFKALYDAPATGVSSSSSTYS 395
DB 1010 ----PYPSTVSGS-----TGSTVSGSTVSGSTASTSGSTSSSTAGSTVS 1053
QY 396 SIVDAVTFPADGFV-----SIVETHAASNGS--MSBQYKSDGEQLSARDLTWSYA 444
DB 1054 GSSASTVTSSTGSGSTGSGSTVSGSTVSGSTITGSGSTESTVTAESTVSGS 1113
QY 445 ALLTANNR-----NSVPAASNGETSASVPGTCAATSAIGTVSSVT--VTSMPSIVAT 496
DB 1114 SVSTVSGNTGSIITGSGTSGSTGSGSTGSESTILESVSIVSVSTGSTITDGSASRSVST 1173
QY 497 GGTATTATPFG--SGSVTSKTTATASKTSTTRSG 531
DB 1174 VSASTESTVSGSSASIGSTINTPDSTESTISGSTITG 1210

RESULT 21
ID 007070 PRELIMINARY; PRT; 190 AA.
AC 007070;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Intracellular glucanase (Fragment).
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPX101-1C;
RA Jimenez A.;
RL Submitted (APR-1989) to the EMBL/GenBank/DBJ databases.

```

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPX101-1C;
 RX MEDLINE=89031230; PubMed=3141213;
 RA Paro U.M., Ianez E., Zalacain M., Clares M.G., Jimenez A.;
 RT "Similar short elements in the 5' regions of the STR2 and SGA genes
 from *Saccharomyces cerevisiae*.";
 RL FEBS Lett. 239:179-184(1988).
 DR EMBL; X13858; CAA32071.1;-.
 DR HSP; P08017; IAXY.
 DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
 DR GO; GO:0005976; P:polysaccharide metabolism; IEA.
 DR InterPro; IPR000165; Glyco_hydro_15.
 DR InterPro; IPR008928; Glyco_trans_6np.
 DR Pfam; PF00723; Glyco_hydro_15; 1.
 DR PRINTS; PR00736; GLHTRLASE15.
 FT NON_TER 190 190
 SQ SEQUENCE 190 AA; 20902 MW; 3799938DB287F40E CRC64;

 Query Match 6.1%; Score 169.5; DB 3; Length 190;
 Best Local Similarity 30.8%; Pred. No. 0.0027;
 Matches 44; Conservative 25; Mismatches 59; Indels 15; Gaps 5;

 QY 1 MSFRSLALSGVCTGLAVYISKRTATDSWLSNEATVARTAILNIGADGAWVSGADSGI 60
 DB 56 VQLRDVILNGTVYVD-SNGAMDSALAEELQOKKVSIEKIFENIGPSAVPS-ISPQV 113

 QY 61 VVASPTNDPFFYTTTRDSGLVLTVDLFRNGDTSLSTENYISAQAIYOGISNPSG 120
 DB 114 VIASPFQTHDYFYQVIRDSALTIINSIVS--HSAGPAIETLLQYINVSFHLQSRNN--- 167

 QY 121 DLSSGAG-----LGEPRKRVND 136
 DB 168 TLGAGIGYNDIVYALGHPKRVND 190

 RESULT 22
 08XOP2 PRELIMINARY; PRT; 4106 AA.
 AC 08XOP2;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Putative hemagglutinin/hemolysin-related protein.
 GN RSP1180 OR RS05070.
 OS *Ralstonia solanacearum* (Pseudomonas solanacearum).
 OC Ralstonia solanacearum.
 OG Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 NC NCB1_TaxID=305;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlat M., Billault A., Brotier P., Camus J.C., Catolico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lave M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Signer P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weisenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*.";
 RL Nature 415:497-502(2002).
 DR EMBL; AL646083; CAD18331.1;-.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0015070; F:toxin activity; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0004812; F:tRNA ligase activity; IEA.
 DR GO; GO:0006418; P:amino acid activation; IEA.
 DR GO; GO:0009405; P:pathogenesis; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR001343; Hemlysn_Ca_bind.
 DR InterPro; IPR000847; HTH_LYER.
 DR InterPro; IPR003995; RTX.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00353; hemolysincbind; 4.
 DR PRINTS; PR00313; CABDNCRPT.
 DR PRINTS; PR01488; RTXTOXINA.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 12.
 DR PROSITE; PS00044; HTH_LYER_FAMILY; 1.
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 4106 AA; 395091 MW; 6338E46766709B46 CRC64;

 Query Match 6.1%; Score 168; DB 16; Length 4106;
 Best Local Similarity 21.6%; Pred. No. 0.23;
 Matches 137; Conservative 70; Mismatches 216; Indels 210; Gaps 29;

 QY 25 ATLDWSLSNEATVARTAILNIGADGAWVSGADSGIVASPSPTNDPFFYTTTRDSGLV 84
 DB 1679 ATAAFTLVDTAAAPATPIIGTVTDVAPVPGA---ITAGGSTNDATPTLTGAFA--- 1730

 QY 85 KTLVDFRNGDPSLSTIE-----NYISAQAIYOGISNPSGSLSSGAGLSP-----K 132
 DB 1731 NSTVSIIF--DGTTLGTTADSGNWSYPTTALTDGSHSLATATDAAGNVSPTSAPT 1788

 QY 133 FNVDETA-----YTG---SMGRPQRDGPALRATA-----MIG 161
 DB 1789 LVYDTAAPATPVISANTDAVAPVTGTTIGSGSTNDAAPTLTGAENSTTNVFDGTTLLG 1848

 QY 162 -----FGQW-----LLDNGY--TSTATDIWPLVRNDLSYVAOYMNQGYDLMEVN 206
 DB 1849 TTADALGSMFTPTTALTDGSHSLATATD-----AAGNVS-----T 1886

 QY 207 GGSFFITAVQHRLVSGSAFATVSGSSCMQSOAPEILCYLOSFWTGSPI----- 257
 DB 1887 ASSAFLLT-----VDIAPAPVIGTVI--DAVAV-----TGTTAGGSTNDAN 1929

 QY 258 -----LANFDSRSRGDANTLLGSI-----HTFDEPAACDS--TFQCSPPALANHK 303
 DB 1930 PTLTGIAEGNSTVYSITDGTTLGTTADASGNWSYPTTALTDGSHSLATADAAGN-- 1987

 QY 304 EYVDSFRSIVTINDGLDSEBAVAVGRYPED----- 333
 DB 1988 --VSSASSAFTLVDPDAAAPAVIGTVDDVAPITGTVAAGSTNDTPTLATAEANT 2045

 QY 334 -TYNNGNPFELCTLAABQLYDALYWDKQGLEVTVDULDFKALYSDAATYSSSSS 392
 DB 2046 INVFDGTTLLGTTA-----DASGNWFTPTSPPLTDG--HSFTATATDA--GNVGTASS 2097

 QY 393 TVSSIVDAVKTADGFSIVETHTAAGNSWSEQYDKSDGQLSARDLWLSYALLTANNR 452
 DB 2098 APTLVDTDAPTIPVISTVYDVAIPVGA-----ITAGGS 2132

 QY 453 RNSVVPASWGETSASSVPQCATSAIGYSSVTYTSW---PSIVATGCT---TTATP-P 505
 DB 2133 TNDAMVLVGTAEANSTISVFDGTTLLGTTADASGNWFTPTTALTDGSHSLATATDP 2192

 QY 506 TSGGSVTS-----TSKTTATASKTSTTTSGMS 533
 DB 2193 AGNVSTTSAFALTVDITAPAAPVITTVDAVS 2225

 RESULT 23
 09BITO PRELIMINARY; PRT; 2016 AA.
 AC 09BITO;
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Fibroin 3 (Fragment).
 OS *Plectreurys tristis* (Spider).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Araneomorphae; Haplogynae; Plectreuridae; Plectreurys.
 NC NCB1_TaxID=33319;

```

RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21179804; PubMed=11283372;
RA Gately J., Hayashi C., Mottink D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
RT Fibroin Sequences";
RL Science 291:2603-2605(2001).
DR EMBL; AF350283; AAK30612.1; -.
FT NON TER 1
SQ SEQUENCE 2016 AA; 196162 MW; F53A47DB185826F1 CRC64;

Query Match 5.9%; Score 164; DB 5; Length 2016;
Best Local Similarity 22.0%; Pred. No. 0.16;
Matches 126; Conservative 75; Mismatches 237; Indels 134; Gaps 24;

QY 35 ATVARTLNNIGADGAWGADSGIYVASTPSTNPNY--FYTWTRDSGLVLTVDLFR 92
D 994 AMARATLTI--IGPDSGTALBSAVVA---VSNVYASAYARALAPAINV--LTN 1046
QY 93 NG-----DTSILSTIENY-----ISAQAIYQGI-----SNPSGDLSSGAG 127
D 1047 NGIFASASEALYLPAPMIAISLHAFGKSFSESSAFALANSISPETAITSQSSSVSAGAS 1106
QY 128 LGPKFNVDTATYTG-----WGRPDGPALRATPMIGFGQMLDNGTST 174
D 1107 SGQSYDTSSVSSASAEATESSVPTDYQATQIESAAAAA-----SSS 1153
QY 175 ATDIWPLVNDISYVVAQ--YWNQGYDLMEEVNGSFFTIYVQHRALV-----GSA 225
D 1154 AYSQSFSESSASAAASAESEQSYDISSDLSASATIAAASASAYEQFSASGSS 1213
QY 226 FATAVGS-----SCSWCDQAPEILCYQSFWTGSHLANFDSRSGKDNATLL 274
D 1214 AAAAASQONSVDYDALYSASAAASAAAYELFESDAS-----SSSAVAVASQ 1267
QY 275 GSHTPEPAACDSTFGPCSPALANHKRVDSFRITLNLGDSSEAVAVGRYEDT 334
D 1268 GSYDT-----SSDFSSASAAAA---AASAYEKKFL--DASSSSAAAAASQSS 1313
QY 335 YVNGNPFLLCTLAABQLYDALYQ--WDKQSLLEV-----DVSLDFKALYSD 381
D 1314 YESSDVLVSASAAAAASAYQSFLDASSSSNAATTSKROSSYDTSSDFSSASIAA 1373
QY 382 AATGYSSSSSTYSIVDAVKTFADGFSIVETHAASNGSMSEQDKSDQESARDLTW 441
D 1374 AA---AASASYES-----QFSD--ASSSSNAAAASQSSYDTS--SDLVASAGASA 1420
QY 442 SYAALLTANNRRNSVPAWGENS-----ASSVPGCAATSAIGTYS--VVTSMPT 493
D 1421 YESQFLDASSSSNAATTSQSSYDTSDFSSASIAAAAAASASYESQFSDASSSNA 1480
QY 494 VATGTTTATPTGSGSVTSTKTTATVASKTS 525
D 1481 AAAAASQSSYDTSDDLVSASAAAAAASAS 1512

RESULT 24
086AK1
ID 086AK1 PRELIMINARY; PRT; 457 AA.
AC 086AK1;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Similar to delayed anaerobic Gene; Dandp.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyocellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szatranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,

```

```

RA Tunggal B., Cox E., Quail M.A., Plattner M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RA Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116984; AA051397.1; -.
SQ SEQUENCE 457 AA; 48301 MW; 16FA6DB74641AAD CRC64;

Query Match 5.9%; Score 163.5; DB 5; Length 457;
Best Local Similarity 20.8%; Pred. No. 0.022;
Matches 96; Conservative 52; Mismatches 130; Indels 183; Gaps 17;

QY 155 RATAMIGFGQWLNDNGY-----TSTATDIWPLVRND--LSYVAQ-----YWNQTS 198
D 53 KDTFMGY-----SGQNLPFTSTLPPTPPYNYQTLGLARSHSSEMGIGYSHNS 106
QY 199 YDLMEEVNGSFFTIYVQHRALVESAFATAVGSSCSWCDs-----Q 240
D 107 YD-----GTSWSTRPDSFVAGCOSAGNSIAENIAAGVSGIGT 144
QY 241 APEILC-----YLSQFWGSHLANFDSRSG 267
D 145 TSQLLCDVAGVCVVDTS DAGRKINMGYKIVGIFPNYSSQYGNWTDQFDNGICG 204
QY 268 KDANTLLGSIHTPEPAACDSTFGPCSPALANHKRVDSFSSI-----YTLN 316
D 205 LPTTPYISGPHT-----FGANAPRYIYSSTSDPAAKVFIIFENGTSIDMTLF 255
QY 317 DGLSDEAVAVGRYPEDTYN--GNPFLLCTLAABQLYDALYQDKQSLVTVSLDF 374
D 256 GGTANS-----GIYTSYVQVCAKYFVSYSASGK-----TYRYPDTGYQLT----- 299
QY 375 FKALYDAAGTYSSTYSIVDAVKTFADGFSIVETHAASNGSMSEQDKSDQESQ 434
D 300 -KA--STCASWVASSSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 341
QY 435 SARDLTWSYAALLTANNRRNSVPAWGENSASVGTCAATSAIGYSSVTVTSMPSIV 494
D 342 -----TSTTSKPTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 373
QY 495 ATGTTTATPTGSGSVTSTKTTATVASKTST--TTRSGMS 533
D 374 ST--TSTTSKPTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 412

RESULT 25
08PD38
ID 08PD38 PRELIMINARY; PRT; 1333 AA.
AC 08PD38;
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Ice nucleation protein.
OS XCC0507.
GN Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro U.A., Reinach F.C., Farah C.S., Furian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Canavan F., Cardozo J., Chambergo F., Clapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria U.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

```

RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Melandri J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezsa R.I.D.,
RA Trindade dos Santos M., Trufi D., Teal S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.,
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL: AE012148; MAM39823.1; -
DR GO: GO:0009279; C:external outer membrane (sensu Gram-negative. . .; IEA.
DR Interpro: IPR000258; Ice_nucleatn.
DR Pfam: PF00818; Ice_nucleation; 68.
DR PRINTS: PR00327; ICENUCLEATN.
DR PROSITE: PS00314; ICE_NUCLEATION; 40.
KM Complete proteome.
SQ SEQUENCE 1333 AA; 131306 MW; 326078458D0E4842 CRC64;

Query Match 5.9%; Score 163; DB 16; Length 1333;
Best Local Similarity 23.6%; Pred. No. 0.1;
Matches 137; Conservative 55; Mismatches 235; Indels 154; Gaps 28;

QY 38 ARTALINIGADGAWGADGIVASPTDNDPYTWRDGLVLTLDLFRNGDTG 97
DB 394 ARTGSLDTAGYGSTAGADSTLIAG-----YGTQTSG-----GDSG 431
QY 98 LUSTIENYISAOIVOGISNPSGDLSSG-----AGLGPKNVDEIAYTGS 143
DB 432 LTA--GYSTQYARKG-----SDLTTGYGSTAGADSTLIAGYSGTQSGESSLTAG 483
QY 144 WGRPQ--PDGPALRA-----TAMIGF-----GOWLNGYSTATDVLWPL 182
DB 484 YGSTQYARKGSLDTAGYGSTAGADSTLIAGYSGTQSGESSLTAGYSTQT----- 537
QY 183 VRNDLSYVAQYVNOT-----GYDLMEVNGSSFFTIAY-QHRAVEGSAFATVAG 231
DB 538 ARKGSDLTLAGYGSTAGGSDTLVAGYSGTQSGDSLTAGYSGTQYARKGSDLTGTYG 597
QY 232 SSGSWCDQAPELCY--LQSFWTGFIILNFDG--SRGKQANLLGSIHFPDEBAC 286
DB 598 ST-STAGGSLTLAGYSGTQSGNASSLTAGYSGTQYARKGSDLTGTYGSTST-----AG 651
QY 287 DDSTFQF--CSPRALNHKEVDSFRSIYTLNDGL-----SDSEAVAVGRYPEDTYN 337
DB 652 ABSTLIAGYSGTQSGESSLTAGYSGTQYARKGSDLTGTYGSTTGA-----DTLLA 706
QY 338 GNPWFLLCTLAALYQMDKQSLVTDVSLDFKALYSDAATGYSSSSSTYSSSI 397
DB 707 G--YGTQYARKGESSLTAGY-----GSTQYARKGSDLTAGYSGTQYARKGSDLTGTYG 758
QY 398 VDAVTKFADGFSIV-----ETHAASNGM-----SEQYKDDGEOLASRD----- 438
DB 759 -----TQYAGFKSLTGYGSTQYARKGSLTLAGYSSSTAGDSSSLIAGYSGTQYARKG 813
QY 439 -LTWGYAALLTANNRNSVVPASWGETSASVPGCAATSAIGTYSVTVTSWPSIVAT 496
DB 814 SLITAGYSGTQYARKG--STLTGTYGSTTGAHDST-----LIAGYSGTQYARKGSLT 866
QY 497 G-GTTTATPT-----GSGSVTSTKTTATATASKTSTTTTREG 531
DB 867 GYGSTQYARKGESSLTAGYSSSWAGPDSLIAGYSGTQYARKG 907

RESULT 26
Q9CH86 PRELIMINARY; PRT; 1063 AA.
AC Q9CH86;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Unknown protein.
GN YIHD OR L10852.

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]_TaxID=1360;
RP SEQUENCE FROM N.A.
RC STRAIN=111403;
RA MEDLINE=2125186; PubMed=11337471;
RX Bolotin A., Wincker P., Mauer S., Jallion O., Malarre K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis 111403."
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006319; AAK04950.1; -
DR PIR: D86731; D86731.
KM Complete proteome.
SQ SEQUENCE 1063 AA; 104547 MW; 5727AD9C69669A1 CRC64;

Query Match 5.8%; Score 161; DB 16; Length 1063;
Best Local Similarity 22.3%; Pred. No. 0.1;
Matches 132; Conservative 90; Mismatches 227; Indels 142; Gaps 24;

QY 9 LSGVCTGLANYSKRA--TLDSWLSNEATVART-----ILNINIGADGAWGADSGI 60
DB 425 LAGLVNNSAIVRGQIAMSQSLPTISNIPLVGTNNVLSPTLNNL-----TGA SLGE 477
QY 61 VVASPTDNDPYTWRDGL-----VLKTLVDFRNGDTSLSTIENYISAOIVOGISN 117
DB 478 VA-----NLTGYSLIDQVNSLGNILISGSTA-LATYEN-----TLQNSLN 518
QY 118 PGSDLSGAG-----LGEPPNVDEIAYTGSWGRPQDGPALRAATAMIGFGWLMD 168
DB 519 SFENLPAGASDLINQVLNQALNNIVESA-TGI-----VNNLPGLG--AIE 563
QY 169 NGTSTATDI--WPLVRNDLSYVAQYVNOTGYDLMEVNGSSFFTIAYQHRALVEGSA 225
DB 564 NGLSNTISQIPNNINFNANLNGITTIIN--SLTPSVGAST-----VNPNSANSNQ 614
QY 226 FATVAGSGSWCDQAPELCYLQSFWTGFIILNFDGSDSEAVAVGRYPEDTYNPNWPL 285
DB 615 SSSSASSSSSSAASSST-----SSSNVSNSTSSNS--EATTSSTSSASSSSSS 661
QY 286 CDDSTFQF--CSPRALNHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNPNWPL 343
DB 662 SESSSASSSSSSSSSVASSSSVDSQSSAGVASSSSSABGSSASSNSSE----- 712
QY 344 CTIAAALYQMDKQSLVTDVSLDFKALYSDAATGYSSSSSTYSSSI-----V 398
DB 713 -SSVASSSSSVDS--SQSSAGVNGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 769
QY 399 DAVTKFADG--FVSIVETHAASNGMSEQYKDDGEOLASRDLTWGYAALLTANNRN 454
DB 770 NSSSSABGSSASSSSSSSSSTSSSSSVGSSQSSSTGVSSSSSSABGSSSHS--SSNSSES 827
QY 455 SVVPASWGETSASVPGCAAT-----SAIGTYSV----- 485
DB 828 SVASSSSSVDSQSSAGVASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 887
QY 486 -TTSWPSIVATGTTTATPTGSGSVTSTKTTATATASKTSTTTTREG 534
DB 888 SSS 938

RESULT 27
Q8VQ99 PRELIMINARY; PRT; 2283 AA.
AC Q8VQ99;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Serine-threonine rich antigen.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcaceae.
OX NCBI_TaxID=1280;

DR GO:0005976; P:polysaccharide metabolism; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR000165; Glyco_hydro_15.
 DR InterPro: IPR008928; Glyco_trans_6hp.
 DR InterPro: IPR000169; SHPoc_acsite.
 DR Pfam: PF00723; Glyco_hydro_15; 1.
 DR PROSITE: PS00639; THIOI. PROTEASE_HIS; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 615 AA; 71427 MW; 8A60F9FACB6A9 CRC64;

Query Match 5.8%; Score 160; DB 17; Length 615;
 Best Local Similarity 21.0%; Pred. No. 0.055;
 Matches 93; Conservative 56; Mismatches 165; Indels 128; Gaps 20;

QY 37 VARTALNNGADGAVSGADSGIVASPTDNPDFTYTRSGVLTLYLDFNNGDT 96
 DB 262 ISLVKKNHVDVGSITASSDYSFV---GLYDSTNYCPRDSALSAHAL-DIAGYGD 316
 QY 97 SLSTSTENYISAOAIVOGISNPSGDLSSGAGLGEPRKFNVDETAYTGSW-----GRP--- 147
 DB 317 AMHG--YQYIS-----EVTBEGFLYHKYNPDKT-LASSMHWIIRGQIYP 360
 QY 148 -QRDGPALRATAMIGFGQWLDNGY-----TSTADIWPLVRNDLSYVAQYWNQT 197
 DB 361 IQDEFALQV-----WAIANHYOIYKIDELIDIFKQFLKRALRFLRYVEDGLPKP 412
 QY 198 GVDLMEVNGSSFFETAVQHRALVEGSAFATAVSSGSCWDS-----QAPELICYLQSF 251
 DB 413 SFLMEERYGHIHYTSTYVAGALVASSELARDIGDEVLASDMLDVAEYKKEVL--RRWV 470
 QY 252 WTGSFFILANFDSRSRSGDANTLLGSIHTFPDEAACDSTFQPCSPALANHKEVD--SF 309
 DB 471 HNGRFI-----RIDENGN-----KDLVIDASMSF-----YFGMDVVRDPI 510
 QY 310 RSIYTLNDGLSDSEAVAAGVPEPTY---YNGNPFELCTLAAEQLYDALYQMDKQSL 365
 DB 511 NTRAIENSIXVGGIA--RYENDMYRKVQPNPMTITLMLAEVLD-LGQREK---- 563
 QY 366 EYTDVSLDFPKALYSDAATGTSYSSSTYSIYDAVTFADGVSIVETHAANGSSEQ 425
 DB 564 -----ALD-----YINAMSRALPSGLLPEQ 584
 QY 426 YDKSGEQLSARDLTWSYALL 447
 DB 585 VDPENFTSTSVPLVMSHAFTI 606

RESULT 30

Q25336 PRELIMINARY; PRT; 888 AA.

AC Q25336; 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
 DE Secreted acid phosphatase 2 (SAP2) precursor.
 GN LMSAP2.
 OS Leishmania mexicana.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5655;
 RN SEQUENCE FROM N.A.
 RC STRAIN=MNYC/BZ/62/M379;
 RX MEDLINE=95237187; PubMed=7720697;
 RA Wiese M., Iig T., Lotsepetch F., Overath P.;
 RT "Ser/Thr-rich repetitive motifs as targets for phosphoglycan
 modifications in Leishmania mexicana secreted acid phosphatase.";
 RL EMBL J. 14:1067-1074(1995).
 DR EMBL: Z46970; CA87091.1; -.
 DR PIR: T46726; T46726.
 DR GO:0003993; F:acid phosphatase activity; IEA.
 DR InterPro: IPR000560; HsAc_phsphtse.
 DR Pfam: PF00328; acid_phosphat_1.
 DR PROSITE: PS00616; HIS_Acid_PHOSPHAT_1; 1.

KW Signal.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 888 MATURE SECRETED ACID PHOSPHATASE 2
 FT (SAP2)
 SQ SEQUENCE 888 AA; 91354 MW; 3C30B56967755505 CRC64;

Query Match 5.8%; Score 160; DB 5; Length 888;
 Best Local Similarity 21.1%; Pred. No. 0.091;
 Matches 116; Conservative 79; Mismatches 257; Indels 172; Gaps 25;

QY 29 SWLSNATATATAILNNT-----GADGAVSG--AD-----SGI 60
 DB 154 SWLDNPAAL---HAAIPVIDEHLSDPAIQSAKDAWVEGLCADYNARTNCVLDMYVAA 210
 QY 61 VVASPTDNP-----DYFTWTRDGLVYKTVLDLPRNGDTSLSTI 102
 DB 211 FEAGRLDNATNIKAVYPGQLQEVNAAMPKRVFSWNHTSKLDL-TQGSASQNLAQTVLANI 269
 QY 103 E-----NYISAOAIVOGISNPSGDLSSGAGLGEPRKFNV-----DEATYTSW 144
 DB 270 NMRLSPSYNNMPOYSADHTVTPLATVPGD--QGETTMRPPFAVTIFVELLDQTLADASGW 327
 QY 145 -----GRPDGPALRATAMIGFGQWLDN-GYTSATDIWPLVRNDLSYVAQY---- 193
 DB 328 YVRLIRGNVVKADGTYYFQESGIRKAYCIDKGNKYLATHTGICPL--NSFRRWVDYSRPA 385
 QY 194 -----WNOTGYD-----LMEVNGS-----SFTIAYQHR---- 218
 DB 386 VADGHCAMTQOTQSNDCPRTIADNKPVRSCMLYHVCPSKACPDYIILSAVDHOCYPG 445
 QY 219 -----ALVEGSAFATAVSSGSCWCD---SQAPELICYLQSFWT-----GSFILANF 261
 DB 446 PDVTNPTSSSSSGCTTSSSEGRATSSDVTTSSEGRATSSDATTSSSEGRATSSSD 505
 QY 262 DSSRSKGDANTLLGSIHTFPDEAACDSTFQPCSPALANHKEVDVFSIYTLNDGLSD 321
 DB 506 ATSSSDATTSSSEGRATSSSEGRATSSDATTSSSEGRATSSDATTSSSEGRAT 563
 QY 322 SEAVANGRPEPTDYNNGNWFELCTLAAEQLYDALYQMDKQ---SLEVTDSLDPFKA 377
 DB 564 SSSDVTTSSEGT-----ATSSDATTSSSEGTSSSDATTSSSEGTAT 609
 QY 378 LYSDAATGTYSSSSSTYSIYDAVTFADGVSIVETHAANGSMEQYD---KSDGEQ 433
 DB 610 TSSDATTSS--SSSEGTSSSDATTSSD---VTTSSSSSEGRATSSSDATTSSSEGR 663
 QY 434 LSARDLTWSYALLTANNRNRSVVPASWG-----ETGASSVGTCAATSAIGTYSVT 486
 DB 664 TTSSDATTSSSEGTSSSDATTSSSEGRATSSDATTSSSEGTSSSRSDATTSSSEG 723
 QY 487 VTSWPSIVATGTTTATPTGSGSVTSTKTNTAKTSTTTRS 530
 DB 724 TATSSDATTSSSEGTSSSDATTSSSEGRATSSDATTSSS 767

RESULT 31

Q990Y4 PRELIMINARY; PRT; 2271 AA.

AC Q990Y4; 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
 DE Hypothetical protein SAV2654.
 GN SAV2654 OR SA2447.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OC Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=158878; 158879;
 RN SEQUENCE FROM N.A.
 RP SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
 RC MEDLINE=2131952; PubMed=11418146;
 RX Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,

RA Cui L., Ogunchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
 RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani U. Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekinaka K., Hirakawa H., Kubara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiratake K.,
 RT Whole genome sequencing of methicillin-resistant *Staphylococcus*
 RT aureus."
 RL Lancet 357:1225-1240(2001).
 DR EMBL: AF003366; BAB58816.1; -
 DR EMBL: AP003138; BAB43752.1; -
 DR PIR: P90073; P90073.
 DR InterPro: IPR008985; ConA_like_1ec_g1.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR008009; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR Pfam: PF03345; He_PIG; 2.
 DR PROSITE: PS00847; GRAM_POS_ANCHORING; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 2271 AA; 227844 MW; 7C2A7040D6C8289D CRC64;

Query Match 5.7%; Score 159; DB 16; Length 2271;

Best Local Similarity 20.4%; Pred. No. 0.38; Mismatches 263; Indels 84; Gaps 17;

Matches 116; Conservative 105; Mismatches 263; Indels 84; Gaps 17;

QY 21 ISKRATLDSWLSNEATVARTAILNINIGADGAWVSGADGIVVASPTDNDPYFTWTRDS 80
 DB 748 VTRNSMDS-VSTSGSTQSGSVSTSKADQSASTSGSIVSTASTSKSTSVLSDS 806
 QY 81 GLVLTIVDFRNGDPTSLTIENYISAQAIQGISNPSGDSGAGLGPKNVDETAY 140
 DB 807 VSASKSLSTSESNVSS--STSTSLVNSQSV--SSSSSGSVSKSTSLSDSISNSNTEK 861
 QY 141 TGSWGRPQRDGPALRATAMIGFCOMLDNGYSTATIIVPL-----VRNDLSVQAQYN 195
 DB 862 SEBLSTSTSD-SIKRTSLSDSLSMSTSGSLSGSLSTISGSSSTASLSDSTSNAI 919
 QY 196 QTCYDLMEEVNGSFFTIAYOHRAIVGSAFATAVGSSCWCDSQADEI-----LC 246
 DB 920 STSTSLSEASSTDSISIS-----NSIANSQASASTSKSDQSTISLSTSDSKSMS 970
 QY 247 YLQSF-----WTCGFLA-----NPDSSRSGKANTLLGSIHFDEACDDSTFQ 292
 DB 971 TSESLSDSTSGSVSGSLSLIASQSVSTSDSMSTSEIYSDISITSGSLASDSKMS 1030
 QY 293 PCSPRALANKEVDSFRSIYTLNDGSDSEAVAGRPDYNGNPMFLCTIAAEQL 352
 DB 1031 VSSSMSTSGSGSTSESLSDSGSTSD--SDSKSL-----STSGSGSTSTSTSTASVKT 1083
 QY 353 YDALYQDKQGLSEVTVSLDFPKALYSDATGTYSSTSSYSIVDAVTFADGFVSIV 412
 DB 1084 SES---QSTSGSMASQSDSMSTSTSPSDSTSDSKASTASSEISQASASTSGSVSTS 1140
 QY 413 ETHAASNG-----SMSEQYDKSD--GEQLSARLUTSYAALLTANN 451
 DB 1141 TSLSTNSEKTSQSVSDSTSLSTSESDISSESTSTSDISEAISASST--SISLSRNS 1198
 QY 452 RRNSVVAWNGETASASVPGTCATSAIGYSSVTVVSWPISVATGTTTAA---TPTGS 508
 DB 1199 TSSSESGASAFLESESLSESTSESTSESVSSTSESTSLSDSTSESGSTSTSLNSITGS 1258
 QY 509 GSV---TSTSKTATATAKTSTTTTRSGMS 533
 DB 1259 ASISTSTSISESTSTFKSESVSTSLSMS 1286

RESULT 32

Q9URU4

PRELIMINARY; PRT; 1283 AA.

Q9URU4

Q9USQ3; PRELIMINARY; PRT; 1283 AA.

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Hypothetical protein (SPBC884.07c protein) (SPBC1289.15 protein).

GN SPBC1289.15 OR SPBC884.07c.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

NC NCB1_faxid=4896;

RN NCB1_faxid=4896;

RP SEQUENCE FROM N.A.

RC STRAIN=972h-1.

RA Wood V., Gwilliam R., Rajandream M.A., Barrell B.G., Oliver K., Harris D.,

Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=972h-1.

RA MEDLINE=21848401; Pubmed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Barrell B.G., Oliver K., Lynne M., Stewart A.,

Scouras J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,

Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.,

Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,

Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagals K.,

James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

Mooney P., Mouton S., Mungall K., Murphy L., Niblett D., Odell C.,

Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,

Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,

Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,

Woodward J., Volkart G., Aert R., Robben J., Grympey B.,

Waltens I., Vanstreels E., Rieger M., Schaefer M., Muller-Auer S.,

Gabel C., Fuchs M., Dusterhoft A., Fritze C., Holzer E., Moestl D.,

Halbert H., Borzym K., Langer I., Beck A., Leirach H., Reinhard R.,

Pohl T.M., Eger P., Zimmermann W., Wedler H., Wandt R., Purnelle B.,

Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,

Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,

Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

Domiguez A., Revela T.L., Moreno S., Armstrong J., Forsburg S.L.,

Ra Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

Ra Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;

"The genome sequence of *Schizosaccharomyces pombe*."

RL Nature 415:871-880(2002).

DR EMBL: AL035675; CAB38695.1; -

DR EMBL: AL021815; CAB6153.2; -

DR GeneDB Spombe; SPBC1289.15; -

DR GO: GO:0016043; P:cell organization and biogenesis; ISS.

DR InterPro: IPR000923; BlueCu 1.

DR InterPro: IPR003324; SGXSG.

DR Pfam: PF02359; SGXSG; 24.

DR PROSITE: PS00196; COPPER_BLUE; 1.

KM Hypothetical protein.

SQ SEQUENCE 1283 AA; 129532 MW; 0B8057CA41C82796 CRC64;

Query Match 5.7%; Score 158; DB 3; Length 1283;

Best Local Similarity 21.0%; Pred. No. 0.2; Mismatches 244; Indels 124; Gaps 24;

Matches 118; Conservative 76; Mismatches 244; Indels 124; Gaps 24;

QY 5 SLIALSGLVCTGLANVSKRATLDSWLSNEA-TVARTAILNINIGADGAW-VSGADSGIV- 61
 DB 180 SIIPHSITSLTSTVSVINDTASLSTKTSPTNGITITETIVSGSVGYSTTFPASGTTSTVE 239
 QY 62 VASPTDNDPYFTWTRDGLVLTIVDFRNGDPTSLTIENYISAQAIQGISNPSGD 121
 DB 240 VVEPT-----ACTITETIV-----SGSVGYSTTFPANGTSGTVEVEPTAG- 281
 QY 122 LSSGAGLGPKNVDETATYNGRQRPDGPALRATAMIGFCOMLDNGYSTATDITVMP 181
 DB 282 -----TITETIVSGS-----VGYSTTFPANGTSGTVEVEPT 313
 QY 182 LVENDLSVYAQYNGQCYDLMEEVNGSFFTIAYOH-----RALVEGSAFAT-----AV 230
 DB 314 TAGTITETIVS--GSVGYSTTFPANGTSGTVEVEPTACTVETITIVSGSVGYSTTFPAS 371

QY 231 GSSCSNCDQAPFILLYLOSFWTGS-FILANPDSSRSGKQANTLLGSIHTFDEAACDS 289
 DB 372 GTTSGTVEVEPEPAGTITETITIVSGSKAFSTFPAN-----GTTSGTVEVEPEPAGTITK 425
 QY 290 TTPQPCPPALANKKEVVDSPRSIYTLNDGSDSEAV---AVGYPEPTYNKNPMLCTL 346
 DB 426 TIVSGS-----KFTSTFPANGTTTGTVEVEPEPAGTITE-TIVSGSVGYTSFP 473
 QY 347 AAEEQLYDALYQMDKQSLVETDVSIDFPKALYSDAATGTYSSSSSTYSIVDAVTFPAD 406
 DB 474 PA-----NOTTGTVEVEPEPT-----AGTITETIVSGSKTFTSTFPASGT-TS 515
 QY 407 GFVSIYTHAASNGSMEQY---DKSDGEQLSARDLTWSYALL--TANNRRNSVVPASW 461
 DB 516 GTVEVEPEPTA---GTITETIVSGSKAFSTFPANGTTTGTVEVEPEPAGTITETIVSGSV 572
 QY 462 GFTSASVETCATSAIGTSSVYTT-----SWPVIATGCTTTATPTGSG 509
 DB 573 GYTSFPASGTTSGTVEVEPEPTAGTETIVSGSVGYTSFPASGTTSGTVEVEPEPT-AG 631
 QY 510 SVTSTKTATATASKTSTTTRSG 531
 DB 632 TITETI-VSGSVGYTSITFPASG 652

RESULT 33

OBIFX6 PRELIMINARY; PRT; 2232 AA.

AC 081FX6;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Hypothetical protein K06A9.1b.
 GN K06A9.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Geisel C., Gattung S.;
 RT "The sequence of C. elegans cosmid K06A9.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U80846; AAC70890.1; -.
 DR PIR; T34434; T34434.
 DR WormPep; K06A9.1b; CB19525.
 KW Hypothetical protein.
 SQ SEQUENCE 2232 AA; 213843 MW; 08D69FA638E14C8 CRC64;

Query Match 5.6%; Score 156.5; DB 5; Length 2232;

Best Local Similarity 23.4%; Pred. No. 0.54;

Matches 113; Conservative 70; Mismatches 228; Indels 71; Gaps 18;

QY 80 SGVLVTLVDLFPNGTSLSTIENTISAGAIYQGISNPSGDSLSSGAGLGEPEKFNVDETA 139
 DB 1440 SSGTQHTTWSKASGST---SPSTNSQTSVTWSSSTSGVSTSSASSTQPMQSTSGS 1496
 QY 140 YTGSMWRPORDGPALPATAIMIGQWLLDNGYTSATDIWPLVRNDLSVVAQYMNQGY 199

DB 1497 SAGSTVASTAPASSTAPASSTGT-----MSSTSSGTGSGTI-SESSSTTASASQTC- 1548
 QY 200 DLMEVNGSSPFTIAVQ-----HRAIVEGAPATAVSSCSW--CDQAPFILLCYL 248
 DB 1549 ---STVTMSSSTSSGVTSSASTQPMSTQSSASGTVASSTALVSTVPSSTGM 1605
 QY 249 QSFMTGSFILANFDSSRSGKQANTLLGSIHTFDEAACDSSTFPQCPSPALANKKEVDS 308
 DB 1606 GSTTSGGT-VGSTISESTTASASQTSVTWSSSTSGVSTSSASTQPMSTQSSA 1664
 QY 309 FRSIYTLNDGSDSEAVANGRPEDTYVGNP---WFLCTLAAEQLYDALYQMDK----Q 362
 DB 1665 GSTVASSTTGL-----VSTSTVPSSTGTWSSSTSGVSTISBSSTPASASQTSSTVTM 1719
 QY 363 GSLEVTDSLDFPKALYSDAATG---TVSSSSSTYSIVDAVKTADGFVSIYTHAAS 418
 DB 1720 GSSSTSGVST-----SSASQGPOMSTSGSSASGTVASSTAPAS-----STAPSS 1767
 QY 419 NSMSQYDKSDGEQLSARDLTWSYALLTANNRRNSVVPASW-----ETSASVPGT 472
 DB 1768 TGTMSSTSGTVGSTMQ-----SSTAASTSHTSITVLLGSSSTSSNQMTSGSVGS 1822
 QY 473 CAATSAIGTSSVYTTWSPVIATGCTTTATPTGSGVTSKTTATA-SKTSTTTRSG 531
 DB 1823 TVASSTAGLVSTV---PSTGTWGSTSGT---VGSTISESTTASASQTSVTW 1876
 QY 532 MS 533
 DB 1877 SS 1878

RESULT 34

093N36 PRELIMINARY; PRT; 1306 AA.

AC 093N36;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Ice protein.
 GN ICEA.
 OS Pantoea ananatis (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxId=553;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=110;
 RA Tang C.R., Sun F.Z., Zhao T.C.;
 RT "Cloning and characterization of an ice gene from Pantoea ananatis
 RT 110.";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF387802; AA70465.1; -.
 DR GO; GO:0009279; C:external outer membrane (sensu Gram-negative. .; IEA.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000258; Ice_nucleatn.
 DR InterPro; IPR000169; SHprot_acsite.
 DR Pfam; PF00818; Ice_nucleatn; 68.
 DR PRINTS; PR00327; ICENUCLEATN.
 DR PROSITE; PS00314; ICE_NUCLEATION; 47.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 2.
 SQ SEQUENCE 1306 AA; 125852 MW; BC1FF266BDF1606F CRC64;

Query Match 5.6%; Score 156; DB 2; Length 1306;

Best Local Similarity 22.7%; Pred. No. 0.28;

Matches 127; Conservative 66; Mismatches 229; Indels 138; Gaps 26;

QY 47 GADGAVSGADGIVVASTNDPQFYWTDSGLVLTVDLFPN-----GPTSL 99
 DB 299 GADSSLIAGSGTQTAGESSTGTAGSTGTAKG-----SDLTAGVSGTGTAGDS-- 350
 QY 100 STIENTISAQAIYQGISNPSG-----DLSG-----AGLGEPEKFN 134

Db 351 SLIAGYSTAGEDSLTAGYGTOTQKGSDLTAGYGTGAGADSLIAGYGTOTA 410
QY 135 VDETAATYGSWGRPQRDPALRATAMIGFQWLLDNGYTATDIDWPLVNRDLSYVAQY 194
Db 411 GBSSTQAGYGTQ-----TAQKGS-----LTAGYGTGT-----AGDSSLIA--- 450
QY 195 NQGYDMEEVNGSSFFTIAGVORALVEGSAFATVAGSSCWCDQAPILCT--LQSF 252
Db 451 ---GYGTQAGEDSLTAGYGTOTQKGSDLTAGYGTGAGADSLIAGYGTOTAG 507
QY 253 TGSFIIANFDSRSRGKANTLL---GSIHFPDPAACDDSTFQPCSPRALANKEVVD 308
Db 508 YGSLTLTAGYGTOTQAGNESDLITGYGTST---AGANSSLIAGYGTOTASINSVLTAG 563
QY 309 FRSIYTLNDGI-----SDSEAVA-VGRYPEDTYNG-NPWFCTLAABEOLY 353
Db 564 YGSLTQTRREGSDLTAGYGTGAGSDSLIAGYGTOTASINSLTAGYGTOTQARQ-- 621
QY 354 DALYQMDKQSLVETDVSLPFKALYSDAATGYSSSSSTYSIVDAVKTFADGFVIVE 413
Db 622 -----SVLTGYGTSTAGADSLIAGYGTOTAGYNSTLITL-----GYGS--- 662
QY 414 THAASNGS-MSEQYDKSD-----GEOLSA---RDLTWSYALILTANRRNSVY 457
Db 663 TQTRQESDLTAGYGTSTAGADSLIAGYGTOTASINSLTAGYGTOTA--REGSVL 720
QY 458 PASWGETSASSVPQTCATSAIGTSSVTVTWSPIVATG-GTTTGA-----TPYSGSV 511
Db 721 TTGYGTSTAG-----ADSLIAGYGTOTAGYNSTLITAGYGTQTRQERSDLTAGYGT 775
QY 512 TSTSKTATASKTSTTRSG 531
Db 776 STAGADSLIAGYGTOTAG 795

RESULT 35

Q8NUJ3
ID Q8NUJ3 PRELIMINARY; PRT; 2275 AA.
AC Q8NUJ3;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE MM2575 protein.
GN MM2575.
OS Staphylococcus aureus (strain MM2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
NCBI_TaxId=196620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "genome and virulence determinants of high virulence community-
acquired MRSA."
RL Lancet 359:1819-1827 (2002).
DR EMBL; AP004831; BAB96440.1; -;
DR InterPro: IPR008985; CONA like lec g1.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR008009; He_PIG.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF05345; He_PIG; 2.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KM Complete proteome.
SQ SEQUENCE 2275 AA; 228407 MW; E676B6BB9E6CE2E CRC64;

Query Match 5.6%; Score 154.5; DB 16; Length 2275;
Best Local Similarity 20.0%; Pred. No. 0.74;
Matches 110; Conservative 106; Mismatches 235; Indels 99; Gaps 16;

QY 17 LANVISKRATLDSLSNE-ATVARTALINNIG-----ADGAWVGASGIVAVSPSTDNP 70
Db 803 LSDSVASKSLSTSESVSSSTSTSLVNSQSVSSMSDSKSKSTSLSDSTSNSSSTKES 862

QY 71 DYFYTWTRDSGLVLTLDVIFRNGDTSLSSTENYISQAIVOGISNPSSGLSGAGLE 130
Db 863 ESISTSTSDS---LRSTSL---SDLSMSSTSGSLSKSOSLSTSESS---STSAISLD 913
QY 131 PRFVDETAATYGSWGRPQRDPALRATAMIGFQWLLDNGYTATDIDWPLVNRDLSYV 190
Db 914 STSNAIST-----ESLSBSASTSDSISINISNOSAST 949
QY 191 AQYMNQGYDMEEVNGSSFFTIAGVORALVEGSAFATVAGSSCWCDQAPILCTLOS 250
Db 950 SKSDSGSTISLSTDSKMSMT-----SESLDSTSTSGSVSGSLIAAGS----- 996
QY 251 FWTGSFIIANFDSRSRGKANTLLGSIHFPDPAACDDSTFQPCSPRALANKEVVD 310
Db 997 -----VSTSTSDMSTSEIVDSISTGSLASDSKMSVSSSMSTISQSGSTSESL 1048
QY 311 STITLNDGLSDSEAVAVGRYPEDTYNGNWFCTLAABEOLYDALYQMDKQSLVETDV 370
Db 1049 DSQSTSD--SDSKSL--STSGSGSTSTSTSTSAVRIS--QSTSGMSASOS 1098
QY 371 SLDPFKALYSDAATGYSSSSSTYSIVDAVKTFADGFVIVETHAASNG----- 420
Db 1099 DSMSTSPEDSTSDSKASATASSEISQASSTSGSVSTSTSTSNERTSTMSDS 1158
QY 421 -----SMSEQYDKSD--GEOLARDLTWSYALILTANRRNSVVPASWGETSASV 469
Db 1159 TSLSTESDSISESTSTSDISEAISASEST--FISISENSTSDSEOSASFLSESL 1216
QY 470 PGTCAATSAIGTSSVTVTWSPIVATGTTTGA-----TPYSGSV---TSTSKTATASK 523
Db 1217 ESTSESTSESVSSSTSESTSLSDSTSEGSTSTSLNSTSGSASISTSTSESTPKS 1276
QY 524 TSTTTRSGMS 533
Db 1277 ESTVSTLSMS 1286

RESULT 36

Q74851
ID Q74851 PRELIMINARY; PRT; 1131 AA.
AC Q74851;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical serine rich protein.
GN SPC18.01C OR SPC17.07C.
OS Schizosaccharomyces pombe (Pisgion yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Hilbert H., Duesterhoeft A., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 44-1131 FROM N.A.
RC STRAIN=972H-;
RA Murphy L., Harris D., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031907; CAA21415.1; -;
DR EMBL; AL031543; CAA20837.1; -;
DR PIR; T41144; T41144.
DR GeneDB_Spombe; SPC18.01C; -;
DR InterPro: IPR005556; SUN.
DR Pfam; PF03856; SUN; 1.
SQ SEQUENCE 1131 AA; 116770 MW; 9A3506989CBEC2DC CRC64;

Query Match 5.5%; Score 153; DB 3; Length 1131;
Best Local Similarity 20.4%; Pred. No. 0.35;
Matches 112; Conservative 85; Mismatches 199; Indels 154; Gaps 22;

Query 32 SNEATVARTALNNIGDAGVAGSGIIVASPSTNDPFIYTWTRDGLVLTLDVLF 91
 Db 560 STALSNLSSTSSASTSYIPSSASSYVANSND-YS-----QTVSIT 606
 QY 92 RNCDDTLLSTIENYISAQAIVQGISNPSGDLSSGAGIGEPKENVND-----ET 138
 Db 607 ASCTTSSTSEI-----VSTPASNSNTGSLNGTSSFNVNSGPFSSQTTPTSSS 654
 QY 139 AYVSGNRPGRDGPALRATAMIGFGWLLDNGYSTATDVLWPLVRNDLSYVAQYNNQTG 158
 Db 655 SLTGSQSLKETSSPAY-VSSTVSTSSVDSSTVNST-----GSSSDSQSPGTT 705
 QY 199 YD-----LMEEV-----NGSFFTIIVQHRALVEGSAFATVGS 232
 Db 706 YSDPTTITSEVSSILSPTSMQSVSRPSSGDAGFNITPISGSGSGTSGYTISS 765
 QY 233 SCGWCDSDQABELLYQSFMTGSPFILANPDSRSKCDANTLGSIH-TPPEACDDSTF 291
 Db 766 NSSQNSASEPQ-----TAFSSSSSATPTTQSSISTSVSSQSMNSYS 810
 QY 292 QPCSPALANHKRVDSF-RSITTLNDGLDSEAVANGRIPEDTYNGNWFCLTAAE 350
 Db 811 SPISNSVTSSTSISSIASSTYSIPIS-----SIAS 844
 QY 351 QLYDA-----LYQMDKQGLEVTDVSLDFKALYSDAATGYSSSSSTYSIVDAVKTFA 405
 Db 845 SFPDAGGFTSYNGTARG-----FSSSF--ALANSSEASDVLSTI-----AKPTF- 890
 QY 406 DGFVSIVETHAANGSNGSEQYDKSDGEQLSARDLWTSYALLTANRRNSVVPASWGETS 465
 Db 891 -----KFSTNSGTSYSPSS-----SRNEGTSYSNITVTS--STLKLSTLSVS 937
 QY 466 -ASSVGTCAATGAI-----TYSSVTVTSWPSIVATGTTTATPPGSGSVTSTKTAT 520
 Db 938 TASSYASASSNTLSTPEKTFSSSSTLS-ESISINTNLTYKPESSLSSTTSGLTS 996
 QY 521 ASKTSTTTS 530
 Db 997 SSTTPSSTRS 1006

RESULT 37
 Q8E473 PRELIMINARY; PRT; 1310 AA.
 ID Q8E473
 AC 08E473
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN GBS1529.
 OS Streptococcus agalactiae (serotype III).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=216495;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NB316 / Serotype III.
 RX MEDLINE=2242508; PubMed=12354221;
 RA Glaeser P., Ruenliok C., Buchrieser C., Chevallier F., Frangeul L.,
 RA Meadek T., Zouine M., Couve E., Lailou L., Poyart C., Trieu-Cuot P.,
 RA Kunst F.;
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
 RT invasive neonatal disease.";
 RT Mol. Microbiol. 45:1499-1513(2002).
 RL EMBL, AL766851; CAD47188.1; -.
 DR Segalistic; gbpl529; -.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR TIGRfam; TIGR01167; LPXG_anchor; 1.
 DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1310 AA; 129397 MM; 9FB7ABBE89B8F155 CRC64;

Query Match 5.5%; Score 153; DB 16; Length 1310;
 Best Local Similarity 20.1%; Pred. No. 0.43;
 Matches 113; Conservative 75; Mismatches 209; Indels 164; Gaps 19;
 QY 25 ATLDMSLNEATVARTALNNIGDAGVAGSGIIVASPSTNDPFIYTWTRDGLV 84
 Db 332 ALKNDLTSEBPGVKLDSSEVLNFAFASVNEGDFVTKLSDNLD-----TQIGITIL 386
 QY 85 KTLVDLFRNDTSLSTIENYISAQAIVQGISNPSGDLSSGAGIGEPKENVND 144
 Db 387 K-----VODIMETQQLATGSPYLTNHI-----TYTW 415
 QY 145 GRPQDGPALRATAMIGFGWLLDNGYSTATDVLWPLVR-----NDLSYV 190
 Db 416 TR-----YASTLNKIKARVMPVPPDRIISKTSDKCF 451
 QY 191 AQYNNQTGYDLMEEVNGSFFTIIVQHRALVEGSAFATVGS SCWCDQA----- 241
 Db 452 ATLNNQVA-SIEERVQYNS-----PSVTEHTNKTIVRSIMKLDDEKQETVITQI 502
 QY 242 -PE-ILCYQSFMTGSPFILANPDSRSKCDANTLGSIH-TPPEACDDSTFQPCSPAL 299
 Db 503 NPEGKEMYPASGLGNLYTIIIGSDGT-SGSPVNLNNEVKILK----- 543
 QY 300 ANHKEVVDSPRSIYT-----LNDG-----LSDSEAVANGRIPEDT 334
 Db 544 TNSKNLTSDMDQYDSEPEFEDVTQSYSTNDGSKITIDMKTNISSTISYVVLVKIPKQS 603
 QY 335 --YVNGNPMFLCTIAAEQLYDALYQMDKQGLEVTDVSLDFKALYSDAATGYSSSSS 392
 Db 604 GVLVS-----TVSDINQYGSKYSGHNTISGDDANAEI--KLSSASASASTSAS 654
 QY 393 TYSIVDAVKTFA DGFVSIVETHAANGSNGSEQYDKSDGEQLSARDLWTSYALLTANRR 452
 Db 655 TSASM-----SASTSASTSASMSASTSASTSASTSASTSASTSASTSASTS 701
 QY 453 RNSVVPASWGETSASSVPGCAATSAIGTYSSTVTSWPSIVATGTTTATPGSGSVT 512
 Db 702 STSA--STASASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAST 759
 QY 513 STSKTTATASKTSTTTRSGWS 533
 Db 760 SASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAST 780

RESULT 38
 Q96011 PRELIMINARY; PRT; 2117 AA.
 ID Q96011
 AC 096011
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN B7A16.095.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fattmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL EMBL, AL513445; CAD11404.1; -.
 DR InterPro; IPR002889; WSC.
 DR Pfam; PF01822; WSC; 5.
 DR SMART; SM00321; WSC; 5.
 KW Hypothetical protein.

SQ SEQUENCE 2117 AA; 216907 MW; 1631362948BEBA32 CRC64;
Query Match 5.5%; Score 152.5; DB 3; Length 2117;
Best Local Similarity 20.8%; Pred. No. 0.9;
Matches 124; Conservative 83; Mismatches 251; Indels 137; Gaps 25;
QY 29 SWLSNEATVARTALINNIGADGAWSGADGIVASSTONDPYFYWTND---SGLVL 84
DB 798 SVLTGSAANSOPIALLEHSTSEGGVNT-----IFCNDPOTCTCFNIDMGTSPLT 848
QY 85 KTLVLDLFRN-GDPSLSTIENYISAQAIYOGISNPSGDLSSGAGL-----G 129
DB 849 TEVTLIRFTNKGSSLM--IDKSKPEMGSVLGAQNPSDDLFEQVNIKKGKSESATLFTFP 906
QY 130 EPEKNVDEATYATGSGWRPORDPALRATAMIGFQWMLDNGYSTARDI--VWPLVNDL 187
DB 907 AAPLNADPIYVSGAWTLNVND-----LTFGVHVNFIGLRAIKVGTLPLDSARF 957
QY 188 SYVAQYNQGYDL-----WEEVNGSFEFTTAVGHRALVEGSAFATA-----VGS 232
DB 958 KYLGCCYRDSANRLLETTQAOFPSPDNDNGK-----CQOYAITNKAAPAGTQYTYECWVGR 1011
QY 233 SCGWCDGQAPELLIC--YL-----QSF--WTGSPILANFDSRSKGDKANTLLGSIHTFDE 283
DB 1012 SIPPASLKVDDYLCNTYICEDKDSQFCGVSIMMMWYDTTGYPENGTILAPARPPASK 1071
QY 284 AACDDSTFOPC-----SPRALANHKEVND-----SFRSI-----YT 314
DB 1072 AVWGDEWAGCRTONSAPRALANDRIYGGSSNTTIESCAQACAGSFQEVENGVCYC 1131
QY 315 LNDGLSDEAVAVGRYEDT---YVNGNPFWFLCTLAABQLYDALYQMDKQGSLEVTDS 371
DB 1132 GNN-LINGSTVA---DEKTCNVVCGDPPELCCGSGRSIVY-----KQKGTIVGNPS 1179
QY 372 LDFPKALYSDAATGYSSSTVSIVDAVKTAFD-----GFYSIVETHAASNGSMSEQ 425
DB 1180 TGWSSS--SGTASGTASATASSTSSATASGTPGNPQSIGYSSIGCTSDAVASSLQ 1237
QY 426 YDKSGEQLSARDLT---WSY-----AALLTANRNRNVSVASGETSA--SSVPG 471
DB 1238 GKNTQSVWGLDDCATYCAQKRYGTEYSACFCGNLLNLAARVLTGRCMLCNGNQOQ 1297
QY 472 TCAATSAIGTYSVTVSWPSIVATGTTTATPTGSGSVTSKTTATASKIST 526
DB 1298 ICGSNGLSIMWQULNPNGTSSSVTASGATOSATASGTASSTSSALATTITSS 1352
RESULT 39
QY 0703X4 PRELIMINARY; PRT; 2014 AA.
AC 0703X4;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN SYNW2303.
OS Synecococcus sp. (strain WH8102).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_Taxid=84588;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22825697; PubMed=12917641;
RA Palenik B., Brahmsha B., Larimer F. W., Land M., Hauser L., Chain P.,
RA Lamerlin J., Regala W., Allen E. E., McCarren J., Paulsen I.,
RA Dufrene A., Partensky F., Webb E. A., Waterbury J.;
RT "The genome of a motile marine Synecococcus."
RL EMBL; BX569695; CAB08818.1;
DR EMBL; BX569695; CAB08818.1;
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 2014 AA; 215624 MW; 28594D70E8965DBC CRC64;
Query Match 5.5%; Score 152; DB 16; Length 2014;
Best Local Similarity 20.2%; Pred. No. 0.91;

Matches 125; Conservative 80; Mismatches 217; Indels 196; Gaps 25;
QY 28 DSWLSNEATVARTALINNIGADGAWSGADGIVASSTONDPYFYWTNDPGLVKT 87
DB 1075 DNIMWSEANVQYTA-----DITDTIGSKPYLDPRNLYL--VQPSGSLTKSA 1118
QY 88 VDLFRNGDLSLSTIENY-----ISQALVQGISNP 118
DB 1119 V-LTESGDARLATESPSYGNFASQAPLAVKDYTTIGGESALYKLLFEQTITGSSDP 1177
QY 119 SGLSSGAGLGEKFNVDVETAYTSGWRPORDPALRATAMIGFG-----QWLLD---- 168
DB 1178 VKRYMT-----LNDQVSMKYDM-----ATTTFEHSHTLEDQRLDND 1218
QY 169 -----NGYSTARDIWPVLV-----NDLSYVAQYNQGYDL-----WE 203
DB 1219 GSVFEVANSNTATADTACAKLRQTSDSLFIKQDGNSAFAVTYADGLRVDDVQETLS 1278
QY 204 EVNGSFEFTTAVO-----HVALVEGSAFATAVSSCGWCDGQAPELLICYSFMTGSFIL 258
DB 1279 DVLSRSTKALAVQKSGDLYLVKETTGLTVDEVE-----TSVVY 1319
QY 259 ANFDSRSRG--KQANTLLGSIHTFDEPAACDSTFOPCSPRALANHKEV---VDSFRSI 312
DB 1320 VYDVTYTAGIFKSSAL-----YRTBAENETWS-----QDITGDGVVSGGT 1363
QY 313 YTLNDGLSDEAVAVGRYEDTYVNGNPFWFLCTLAABQLYDALYQMDKQGSLEVTDSL 372
DB 1364 SSASDAFADAVGTDVQEVVDKFNQNSASDILSVKADG--SNPLISIFVFPKSGAKSVNDL 1422
QY 373 D-----FPKALYSDAATG--TSSSSSTYSIVDAVKTAFDGFVIVE----- 413
DB 1423 TVKQVQSVADPLMAKAVYDGLTAQVQSRRAKTTDSYEAVTGLDQITVTVQAKFGKI 1482
QY 414 -----THAASNGSMSE---QYKSDGE--QLSARDLTWGYAALLTANR 452
DB 1483 QSVSWVLPVGVGADPTIYAKNSVTGEFTNIQYDAETEGQWMBETRT---LTLFRDNG 1539
QY 453 RNSVVPASWGETSASVPGTCAATSAIGTYSVTVSWPSIVATGTTTATPTGSGSVT 512
DB 1540 KND-----ESDDQGITRRPGVVAARSGETSDSTTPPTPSPVPTP--PTPATPTPPTPT 1593
QY 513 STSKTTATASKISTTTRS 530
DB 1594 PTPPTPTPTPATPTPTSPS 1611
RESULT 40
QY 0874K9 PRELIMINARY; PRT; 1383 AA.
AC 0874K9;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Agglutinin-like protein.
GN ALS5.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_Taxid=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=SC5314;
RA Zhao X., Nuesen J. A., Hoyer L. L.;
RT "Analysis of Candida albicans ALS5p function."
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY227440; AAO72529.1;
SQ SEQUENCE 1383 AA; 145639 MW; 70A809E8AD79AC41 CRC64;
Query Match 5.4%; Score 150; DB 3; Length 1383;
Best Local Similarity 20.5%; Pred. No. 0.72;
Matches 113; Conservative 84; Mismatches 197; Indels 158; Gaps 26;

QY 26 TLDSWLSNEATVARTAILNNGADGAMVSGADGIIVASPSDNPDPYFT--WTRDSGLV 83
DB 440 TTQPF--SESTSTTTITNSL-----KGTDS-VIYREP--HNPTVTTEFMS----- 481
QY 84 LKTLVLFNRGDTSLSTIENYISAQAIYVG-----ISNPSGLSSAGIGERKEN 135
DB 482 -----ESVATTEITTTGPGTDSVIIKEPNPTVTITVWMSSEYAT 522
QY 136 DETAYTGSWCR-----PORDPALRATAMIGFGOMLDNGYSTATDIWMP-----LVR 184
DB 523 TEITTTGPGTDSVIIKEPNPTVTITTKF-----W--SESYATTEITITNKPEGTDSVIYR 575
QY 185 ---NDLSVYAQYWNQGYDLMEVNGSSPFTIAVORALVEGSAFATAVGSSCWCDSDQA 241
DB 576 EPHNPVTVTTEFMS--YATTEITTTGPGTDSIVIHDPLESSSTTAISSDSNINSSA 634
QY 242 PELTCLTQSWTGSFILANFDSRSGKANDTLIGSIHTFPEAACDDSTFQCPSPALAN 301
DB 635 QE-----SSSVESQSSIVGLSSSDIPLSSD---MPSSS----- 666
QY 302 HKEVDSFRSIYTLNDGLSDSEBAVAVGRYPEDTYNGNPMFLCTLAABQLYDALYQMDK 361
DB 667 -----STGLTSSSESTVSSYDSSSSSE---LSTFSSSESY----- 700
QY 362 QGSLFVTDVSLDFPKALYSDPATGTYSSSSSTYSIVDAVKTPEADGFSIVETHAASNG- 420
DB 701 ---SSISIDTT-NW-----DSSSDLESTISITWSSSIDA-----QSSQSVQSVNSI 744
QY 421 SNEQYDKSDGEO--ISARDLTVSYALLTANNRRNSVVPASV-----GE--- 463
DB 745 STSQETSSSGEBSNTVDILVSDASSILNDSISSYPSSTISLSDPEPHITAGEPOS 804
QY 464 TSASVPGTCATSAIGTYSVVTWSPSIVATGCTTTATPTGSGSVT--STSKTATAS 522
DB 805 RSSSSLASTVETISDLVSLTSDPTSSFDSSSLNSDSSSPFDESDISASSFSFTLVAP 864
QY 523 KSTTTTRSGMSL 534
DB 865 SFLSSSSSSLSL 876

RESULT 41
Q08294
ID 008294 PRELIMINARY; PRT; 967 AA.
AC 008294;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Chromosome XV reading frame ORF YOL155C.
GN YOL155C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1] _____
RP SEQUENCE FROM N.A.
RA Gallion L., Dujon B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2] _____
RP SEQUENCE FROM N.A.
RA Arino J., Casamayor A., Gamo F.J., Gancedo C., Lafuente M.J.,
RA Aldea M., Casas C., Herrero E.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3] _____
RP SEQUENCE FROM N.A.
RA MIPS;
DR EMBL; 274897; CAA39177.1; -
DR PIR; S66852; S66852.
DR SGD; S0005515; YOL155C.
SQ SEQUENCE 967 AA; 94704 MW; 7BFC01EA243A561E CRC64;

Query Match 5.4%; Score 149; DB 3; Length 967;

Best Local Similarity 20.4%; Pred. No. 0.51;
Matches 117; Conservative 55; Mismatches 201; Indels 200; Gaps 20;
QY 10 SGLVCTGLANVSKRATLDSWLSNEATVARTAILNNGADGAMVSG---ADSGIV----- 61
DB 434 SGLISLSESTKGEVTFSPYNSGAFSFSNAILN-----GGSVSLQRRAESGVNNGE 488
QY 62 -----VASPSTDN-----PDYFYWT---RDSGLVLTVDLFRNG 94
DB 489 INLENGSTVWVEPVSQSGTINIISGNLYLHPDRTTGQTGVFKGCVL---AVDPTEN 545
QY 95 DTSL--STIENYISAQAIYOGIS-----NPSGLSSAGIGERKENDET 138
DB 546 TPIPVGYTGENQALITADYATLALSYDANGLVLTATQNSQFSIGTGFSSSGFVNSG 605
QY 139 ATYGSWGRPORDGPALRATAMIGFGOMLDNGYSTATDIWPLVNDLSVYAQYWNQIG 198
DB 606 TRGAVAYVLYNGGVVASATPS-----STSTTSQATN----- 638
QY 199 YDLMEVNGSSPFTIAVORALVEGSAFATAVGSSCWCDSDQAPELICYLQSPWTSFIL 258
DB 639 -----STGSGTSPGASV-----TGSTAS 656
QY 259 ANFDSSRSGKANDTLIGSIHTFPEAACDDSTFQCPSPALANHKEVDSFRSIYTLNDG 318
DB 657 TSFGASVTSSTALISG-----SP-----SVYTTTLT 684
QY 319 LSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLYDALYQMDKQSLFVTDVSLDFPKAL 378
DB 685 YATITSTVAVSCSETTDSNGNVTITITVPCSTTATITISCETGCHVTST----- 736
QY 379 YSDAATGTYSSSSSTYSIVDAVKTPEADGFSIVETHAASMSRQYKSDGEOULARD 438
DB 737 -GTVAETVSSS-----YTYVYTHCNGNCGNTVTSSECEETSAIT 779
QY 439 LT-WSYALLTA---NNRRNSVVPASWGETSASVPGTCATSAIGTYSVVTWSPSIV 494
DB 780 TSPKSYTVTVTHCDNGCNT-----KTVSSEAPETATTVSPRTYTTATVATVQDD-- 830
QY 495 ATGCTTTATPTGSGSVTSTKTATASTTT 527
DB 831 -NGCSTKVTVSEAPKETSETSETSA-APKTYTT 861

RESULT 42
Q05164
ID 005164 PRELIMINARY; PRT; 1001 AA.
AC 005164;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE AOB567, AOF1001, AOE110, AOS264 and AOE130 genes.
GN AOF1001.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1] _____
RP SEQUENCE FROM N.A.
RC STRAIN=FY1679;
RA Gamo F.J., Lafuente M.J., Casamayor A., Aldea M., Casas C., Arino J.,
RA Herrero E., Gancedo C.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X69715; CAA61860.1; -
SQ SEQUENCE 1001 AA; 97673 MW; F11BC4522DCFBDAF CRC64;

Query Match 5.4%; Score 149; DB 3; Length 1001;
Best Local Similarity 20.4%; Pred. No. 0.54;
Matches 117; Conservative 55; Mismatches 201; Indels 200; Gaps 20;
QY 10 SGLVCTGLANVSKRATLDSWLSNEATVARTAILNNGADGAMVSG---ADSGIV----- 61
DB 468 SCDISLISSESTKGEVTFSPYNSGAFSFSNAILN-----GGSVGLQRRAESGVNNGE 522

```

QY 62 -----VASPSTDN-----PDYFTWT--RDSGLVLTIVLDFRRG 94
DB 523 INLENGSTYVVEVPVSGSGTINIISGNLYLHYDPDTFTGTVFVKGEGVL--AVDPRETN 579
QY 95 DTSL--STIENYISQAIVOGIS-----NPSGLSSAGLGEKENVDET 138
DB 580 TTP1PVVGYTGEOIATADVLTALSYDSATGVLTAQNSQFSFGTGFSSSGFVNSEG 639
QY 139 AYTGSWGRPQRDPALRATAMIGFGQWLDNGYSTATDIIWPLVRNDLSYAQYMNQGT 198
DB 640 TPGAGAVYIYANKGVVASSATPS-----STSTTSGATN-----672
QY 199 YDIMEEVNGSSFTTIAVQHRALVEGSAFATAVAGSSCGWCDSQAPBILCYLQSFMTGSFIL 258
DB 673 -----STGSGTSPGASV-----TGSTAS 650
QY 259 ANFDSSASGKDANTLLGSIHTFDEACDDSTPQPCSPRALANKEVVDSPRSIYTLNDG 318
DB 691 TSPGASVYTGSTASTLLSG-----SP-----SVYTTTLT 718
QY 319 LSPSEAVAVGRYPEDTYNGNPNWFLCTLAABQLYDALYQMDKQGLVTDVSLDFKAL 378
DB 719 YATTTSTVAVVSCSTTNSNGNVITTTVCSSTTATITSCDETCGVTTST-----770
QY 379 YSDAATGYSSSSSTYSSIVDAVKTFFADGFVSIETHAASNGSSEQYDKSDGEQLSARD 438
DB 771 -GTVAETVSSKS-----YTVTVTHCDNNNGCNTKTVTSECPRETSAT 813
QY 439 LT-MSYAAALLTA--NNRRSVVPASWGEFSASVPTGCAATSAIGYSSVTVTWSISIV 494
DB 814 TSPKSYTVTVTHCDNNGCNT-----KTVTSEAPATTTTVSPKTYTATVYQCD-- 864
QY 495 ATGGTTTATPTGSGSVTSTSKTATATSKTSTT 527
DB 865 -NGCSTKTVTSEAPKETSETSETSA-APKTYTT 895

```

RESULT 43

```

QY 070737 PRELIMINARY; PRT; 2230 AA.
AC 070737;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE Conserved hypothetical precursor.
GN SYN00985.
OS Synecchococcus sp. (strain WH8102).
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
OX NCBI_TaxID=84588;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825697; PubMed=12917641;
RA Lamerik B., Brahamsa B., Larimer F.W., Land M., Hauser L., Chain P.,
RA Lamerik J., Regala W., Allen E.B., McCarren J., Paulsen I.,
RA Ditrene A., Patensky F., Webb E.A., Waterbury J.;
RT "The genome of a motile marine Synecchococcus";
RL Nature 424:1037-1042(2003).
DR EMBL: BX569691; CAB07500.1; -
KM Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 26 Potential.
SQ SEQUENCE 2230 AA; 226712 MW; C55BA0582586F7E9 CRC64;

```

```

Query March 5.3%; Score 147; DB 16; Length 2230;
Best Local Similarity 21.2%; Pred. No.2.2;
Matches 129; Conservative 90; Mismatches 224; Indels 166; Gaps 31;

```

```

QY 1 MSFRLSLALSG-----IVCTGIANYISKRTATD-----SMLSNE 34
DB 690 LAFSTNTSSISGNSWASGVLTLTSSDSLANKALLESYTYNNTSDNPNTSRIISWVND 749
QY 35 ATYAKTALNNIGA-----DGAWVGADSGIVAVSPIDNPYFTTWTRDSGLVLTIVDL 90

```

```

DB 750 GDTNSSAVTSTISVAADVNDAPVTSASATLASEEDGATP-----IDSSL--TTTDA 799
QY 91 FRNGDTSLSTIEN-----YISAQIVOGISNPSGDLG-SGAG-LGEBKENVDET 138
DB 800 DQSNIESATITISSGQSTEDVLAFLDTSAITGSMWASGVLTLTSSGTLANKALESV 859
QY 139 AYTGSWGRPQRDPALRATAMIGFGQWLDNGYSTATDIIWPLVRNDLSYAQYMNQGT 198
DB 860 TYNNTSDTPNTSNRII-----SMVVDGDTNESA-----VSTISVA-- 897
QY 199 YDIMEEVNGSSFTTIAVQHRALVEGSAFATAVAGSSCGWCDSQAPBILCYLQSFMTG 254
DB 898 -----VNDAPVYAGSATLASEGNG-ATIIDSLITTDADSNIESATVISSGFGSA 950
QY 255 SFLIANFDSS-----RSGKDANTLLGSIHTFDEACDDSTPQPCSPRALANKEVVD 308
DB 951 EDVLAFSDTSALMGWNSATGVLTLGSDSLANKALLESYTYNNTSDNPNTAD- 1004
QY 309 FRSI-YTLNDGLSDSEAV-----AVGRYPEDTYNGNPNWFLCTLAABQLYDALYQMD 360
DB 1005 -RTISWVNDGDTNSSAVTSTITVTAVNNAP--VISGAS--ATLAFSEE--DGATPID 1055
QY 361 KQGLVTDVSLDFKALYSDAATGYSSSSSTYSSIVDAVKTFFADGFVSIETHA-ASN 419
DB 1056 --SSLTITD-----ADNTIESATVISGFGSAEDVLAFLDT 1091
QY 420 GSNSEQYDKSDGE-QLSARDLTWSYALL--TAN-----NNRNSVPASW--GETS 465
DB 1092 SSISGSMWNSATGVLTLGSGTLANKALESYTYNNTSDTPNTSNRII--SMVVDGDTN 1149
QY 466 ASSVPTGCAATSAIGYSSVTVTWSISIVATGTTTATPTGSGSVTSTSKTATASKTS 525
DB 1150 SSASV-----TSTISVAADVNDAP--VYAGSATLASEGAGATIIDSLITTDADS 1198
QY 526 TTRSGMSL 534
DB 1199 NIESATVIS 1207

```

RESULT 44

```

QY 097YV0 PRELIMINARY; PRT; 612 AA.
AC 097YV0;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DE "The complete genome of the crenarchaeon Sulfolobus solfataricus".
GN SSO2473.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=2132296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyer M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Ernaus G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.B., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL: AE006846; MAK42610.1; -.

```

```

DR PIR: C90419; C90419.
DR CO: G0:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
DR GO: G0:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR CO: G0:0005976; P:polysaccharide metabolism; IEA.
DR InterPro: IPR000165; Glyco_hydro_15.
DR InterPro: IPR008928; Glyco_trans_6hp.
DR Pfam: Pf00723; Glyco_hydro_15; 1.

```


KW Hydrolase; Glycosidase; Complete proteome.
SQ SEQUENCE 612 AA; 70938 MW; 235FF5JF614CA96 CRC64;

Query Match 5.3%; Score 146; DB 17; Length 612;
Best Local Similarity 21.0%; Pred. No. 0.43;
Matches 101; Conservative 66; Mismatches 173; Indels 142; Gaps 23;

QY 16 GLANVLSKRLT-----DSWLSN-----BATVARTAIL--NNIGADGAWVSGA 56
D 213 GLSDYRKRTPALFRKVDYWRAMLSKVNDYGEYSILRSLLIIOHQNNGALVAGSL 272
QY 57 DSGIVASSTNDPBYFTWTRDSGLVLTVDLFRNGDTSLSLTENIYSAQAVQGIS 116
D 273 DTDIMKFNEDT---YNYVWHRDAVFCIAL-----ELMGVDRSRQPEFTRRLLF 319
QY 117 NPSGDISGAGLGEPEKFNDETAYTSGMRP-----QRDPALRATM-IGFGQWLL 167
D 320 TINGALFH-----RYTVD--GHFGSTWHPMTLDYLPIDEDETALVYALMWFHFKW-K 369
QY 168 DNGYTSTATDVIWPLVRNDLSVVAQWNO-TG-----YDLMEVNGSSFTIIVQHRALY 221
D 370 DVDFIT---YYPWPKGIADFLVNYRKAQTGLPLSFDMERIGTHFTTTIVAGLR 426
QY 222 EGSAPFATVAGSS--CSWCDSQAPILCYLQSFMTGSFTILANFDSRSRGK-----DAN 271
D 427 AADFAFYRGEDELAKQKETFADQMRNSIDLFWVGDH-YARTIYMEKGQVHKIDETVDS 485
QY 272 TLGSGHTDPEACDSTFQPCSPRALANKEVDSFNSIYTLNGLSDSEAVANGRP 331
D 486 ILIAPLFNVIP---MDSRP-----VKNLETVIEKLSYKRGIV--RYE 523
QY 332 EPTYV---NGNWPFLCTLAALAEQLYDALYQMDKQSLSEVTVSLDFPKALYSDAATGY 387
D 524 GDQYLGAGNNSNWFISTLWS-QVYSLMGKDK----- 556
QY 388 SSSSTSYSSIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYAL 447
D 557 -----AKKIDWVLSKSLPTGVIPQIDND-KVPSVSLAMSHAEI 598
QY 448 TA 449
D 599 RA 600

RESULT 45

Q88789 PRELIMINARY; PRT; 860 AA.
ID Q88789
AC Q88789;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Muramidase (Putative) (EC 3.2.1.17).
GN LP_3093.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxId=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFSL;
RX MEDLINE=22480296; PubMed=12566566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fieze M.W.E.J., Stekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hofier S.M., Nierop Groot M.N., Kerhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Sijzen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFSL.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL: AL935261; CAD65260.1;
DR GO: GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO: GO:0003796; F:lysozyme activity; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR GO: GO:0016998; P:cell wall catabolism; IEA.

DR GO: GO:0009253; P:peptidoglycan catabolism; IEA.
DR InterPro: IPR002053; Glyco_hydro_25.
DR Pfam: PF01183; Glyco_25; 1
KW Glycosidase; Hydrolase; Complete proteome.
SQ SEQUENCE 860 AA; 89631 MW; 4DAC7A298BF90F1C CRC64;

Query Match 5.2%; Score 145; DB 16; Length 860;
Best Local Similarity 20.1%; Pred. No. 0.79;
Matches 139; Conservative 85; Mismatches 280; Indels 186; Gaps 28;

QY 7 LALSLVCTGLANVLSKRLTDSWLSNEATV---ARTAILNIGADGAWVSGADSGIV 62
D 151 VASSTATTTQVSHTS--STAASASTAVYAVKASSTVSAATATKATKAAVAVAY 208
QY 63 ASPTNDPBYFTWT--RDSGLVLTVDLFRNGDTSLSLTENIYSAQAVQGI--SNRS 119
D 209 TAATTSSTD---VMTIGDTRPRVDVDAVSQSTTQSDP--NKLAAGAKTVIVKSTEG 264
QY 120 GDLSSGAGLGEPR-----FNVDEYATYSGMRPQRDGPALRATMIGFGWLLDNG--- 170
D 265 TDYTNPALNQAANKAKALNDVFNHYATP---STADAKSEATNAGF---LVKNVST 318
QY 171 -----YSTA---TDIWPVLRNDLS 188
D 319 KYLLFADMEDSSSYANATANINAEFWSTLDSFGYKNHGVYTSNSYLRDAVTVQGSRV 378
QY 189 YVAQY-WNOTGYDIMEVNG-----SFTIIVQHRALVEGSAFATAG 231
D 379 WRQYPTPSANNLWNTNGAWQFSDTALLPFGSDYTGIVDVIDYNGLTESAGTNTFV 438
QY 232 SSCWCDSQAPILCYLQSFMTGSFTILANFDSRSRGDA-----NTLL----- 274
D 439 TTTSNDTTTSEVT-----TDTSATNTSTSTTKKASGWYTFPKTRAIKSAADSAR 491
QY 275 -----GSIHFDPEAACDSTFQPCSPRALANHK 303
D 492 TVGTYSKGNRYVYNAEITNGETWLRYSLSGSEH-FVXIAAKTTTTPASTSKTYRK 550
QY 304 EYVDSFRSLYT--LNDGLSDSEAVANGRPEDTYY-----NGNWPFLCTLAALAEQLYD 354
D 551 NETGYKFTKTTAIGSVSDSAKTLGTYKGDVYVNAKVTNGETWLRYSLSGAQHVV 610
QY 355 ALYQMDKQSLSEVTVSLDFPKALYSDATGYSSSSST--YSSIVDAVK---TFADGFV 409
D 611 KIS--GAATSTTTKPAATSSSKTV--TKATGYTKFTGTTRAIKGSVVDASAKTLGTYKGP 667
QY 463 ETSASVPEICATATSAIGTYSVTVTSMPSIYATGTTTATPTGSG----- 509
D 728 KTTAIKSSASDAA-STVGTGYKNTVYVNAKVTNGQTLRLRYTSYSGAQHYVKGGAAT 786
QY 510 -SVTSTSKTTATAS---KSTTTRSGML 534
D 787 TTSSASKTTAAAGTYTFTKTINIRTAASL 816

Search completed: June 28, 2004, 07:44:51
Job time : 53 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:40:02 ; Search time 18 Seconds
(without alignments)
1544.748 Million cell updates/sec

Title: US-10-038-723-2
Perfect score: 2771
Sequence: 1 MSFSLILASGLVCTGLANV.....SKTTATASKSTTTTSGMSL 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2742	99.0	640	1	AMVG_ASPNG
2	2620.5	94.6	639	1	AMVG_ASPSH
3	2588.5	93.4	639	1	AMVG_ASPKA
4	1935	69.8	612	1	AMVG_ASPOR
5	1577	56.9	626	1	AMVG_NEUCR
6	1321	47.7	616	1	AMVG_HORRE
7	781.5	28.2	450	1	AMVG_SCHPO
8	725	26.2	604	1	AMVG_RHIOR
9	671.5	24.2	519	1	AMVG_SACFI
10	664.5	24.0	519	1	AMVG_SACFI
11	641.5	23.2	624	1	AMVG_AXAD
12	597	21.5	549	1	AMVG_YEAST
13	522	18.8	767	1	AMVI_SACDI
14	521	18.8	768	1	AMVI_SACDI
15	221.5	8.0	615	1	YGI0_METJA
16	159.5	5.8	1196	1	ICEV_PESX
17	159	5.7	1210	1	ICEN_PESFL
18	158	5.7	670	1	YFSG_SCHPO
19	158	5.7	3178	1	YSH8_CABEL
20	157	5.7	1258	1	ICEN_ERWHE
21	156.5	5.6	540	1	GUX1_ASPAC
22	155.5	5.6	1148	1	ICEN_PESX
23	155	5.6	1132	1	ICEN_PANAN
24	154.5	5.6	1200	1	ICEN_PESY
25	151.5	5.5	1034	1	ICEN_PANAN
26	151.5	5.5	1036	1	MSB2_YEAST
27	151.5	5.5	1331	1	MANB_CALSA
28	150	5.4	1609	1	FIG2_YEAST
29	147	5.3	507	1	YG46_YEAST
30	142	5.1	1419	1	ALAI_CANAL
31	141.5	5.1	644	1	XYND_CELFI
32	141	5.1	1802	1	HKRI_YEAST
33	140.5	5.1	1567	1	ICEN_XANCT

34	138.5	5.0	600	1	SP96_DICDI	P14328 dictyostei
35	138	5.0	797	1	VGUX_HSVB	P28968 equine herp
36	137	4.9	556	1	WSC3_YEAST	Q12215 saccharomyc
37	136	4.9	542	1	CHT2_RHIOL	P29027 rhizopus o1
38	136	4.9	995	1	YI09_YEAST	P40442 saccharomyc
39	135.5	4.9	537	1	GUX1_PENJA	Q06886 pentacillium
40	133.5	4.8	1589	1	PHP_DROME	P39769 drosophila
41	132	4.8	827	1	XAND_XANS2	Q60106 xanthomonas
42	130.5	4.7	1537	1	PL01_YEAST	P32768 saccharomyc
43	130.5	4.7	1746	1	TENA_PIG	Q29116 sus scrofa
44	130	4.7	706	1	PLB2_YEAST	Q03674 saccharomyc
45	129.5	4.7	1367	1	AMVH_YEAST	P08640 saccharomyc
46	129	4.7	1025	1	SLAP_CAUCR	P35828 caulobacter
47	128	4.6	1169	1	YK82_YEAST	P36170 saccharomyc
48	127.5	4.6	776	1	PRV3_YEAST	P26501 pseudomonas
49	127.5	4.6	881	1	PRV3_YEAST	P47033 saccharomyc
50	127.5	4.6	1322	1	YAG3_YEAST	P39712 saccharomyc
51	127	4.6	497	1	FLIC_ECOLI	P04949 escherichia
52	127	4.6	540	1	CH11_RHIOL	P29026 rhizopus o1
53	127	4.6	1140	1	YM96_YEAST	Q04893 saccharomyc
54	126.5	4.6	776	1	ISOA_PSEAY	P10342 pseudomonas
55	126.5	4.6	1861	1	APU_THETU	P38536 t amylopull
56	126	4.5	567	1	CH13_CANAL	P40954 candida alb
57	126	4.5	1161	1	DAN4_YEAST	P47179 saccharomyc
58	125	4.5	525	1	NU62_RAT	P17955 rattus norv
59	125	4.5	962	1	GUIN_PSEFL	P10476 pseudomonas
60	124	4.5	562	1	CHT_YEAST	P29029 saccharomyc
61	124	4.5	1616	1	SLAP_BACCI	P35824 bacillus c1
62	123.5	4.5	5703	1	MUS8_HUMAN	Q9hce8 homo sapien
63	123.5	4.5	503	1	WSC2_YEAST	P53832 saccharomyc
64	123.5	4.5	1075	1	FLOS_YEAST	P38894 saccharomyc
65	122.5	4.4	1382	1	YRF4_YEAST	Q13559 saccharomyc
66	122.5	4.4	1796	1	YRPI_YEAST	P24088 saccharomyc
67	122	4.4	510	1	YD10_SCHPO	O13692 schizosacch
68	122	4.4	699	1	CH11_BACCI	P20533 bacillus c1
69	122	4.4	1041	1	EGT2_YEAST	P24852 saccharomyc
70	121.5	4.4	3584	1	NBEA_DROME	Q9w4e2 drosophila
71	120.5	4.3	1374	1	YMN3_YEAST	Q03099 saccharomyc
72	120.5	4.3	1581	1	YR2_YEAST	P40105 saccharomyc
73	120.5	4.3	1859	1	YR3_YEAST	P53345 saccharomyc
74	120.5	4.3	1859	1	YR6_YEAST	P53819 saccharomyc
75	119.5	4.3	1090	1	GUXB_CELFI	P50899 cellulomona
76	119	4.3	542	1	YLM2_SCHPO	Q9p378 schizosacch
77	119	4.3	693	1	TGM3_MOUSE	Q08189 mus musculu
78	118	4.3	1260	1	ALSI_CANAL	P46590 candida alb
79	117.5	4.2	559	1	GAST_YEAST	P22146 saccharomyc
80	117.5	4.2	1758	1	YIR1_YEAST	P40434 saccharomyc
81	116.5	4.2	1758	1	YJMS_YEAST	P40889 saccharomyc
82	116.5	4.2	542	1	SCWB_YEAST	P53189 saccharomyc
83	116.5	4.2	1045	1	GUIN_CELFI	P26325 cellulomona
84	116	4.2	605	1	WSC4_YEAST	P38739 saccharomyc
85	116	4.2	1005	1	Y456_CHLTR	O84462 schizosacch
86	115.5	4.2	1953	1	BIGH_SALTY	P25927 salmonella
87	115	4.2	412	1	H150_YEAST	P32478 saccharomyc
88	115	4.2	596	1	MKCT_YEAST	P53379 saccharomyc
89	114.5	4.1	569	1	YAP3_YEAST	P32329 saccharomyc
90	114.5	4.1	649	1	YAP3_SCHPO	Q10211 schizosacch
91	114.5	4.1	841	1	YAGX_ECOLI	P77802 escherichia
92	114.5	4.1	914	1	GUX2_CLOSR	P50900 clostridium
93	113.5	4.1	373	1	YLUZ_PICAN	P34735 picia angu
94	113.5	4.1	526	1	NU62_MOUSE	Q63680 mus musculu
95	113.5	4.1	755	1	Y572_CHLPL	O92791 chlamydia p
96	113	4.1	550	1	FLIC_SHIFL	O08860 shigella fl
97	113	4.1	1742	1	GUIN_CALSA	P22534 caldocellum
98	113	4.1	2432	1	Y43R_IRV6	P18305 chilo iride
99	112.5	4.1	420	1	SUN4_YEAST	P53768 saccharomyc
100	112.5	4.1	549	1	EPD1_CANNA	P56092 candida mal

RESULT 1

ALIGNMENTS

AMYG ASPNG STANDARD; PRT; 640 AA.
ID AMYG ASPNG Q92201; Q99179;
AC F04064; Q92201; Q99179;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1986 (Rel. 03, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucoamylase (Gl. 41, Last annotation update) (Glucoan 1,4-alpha-glucosylase) (1,4-alpha-D-glucan glucosylase).
GN GLAA.
OS Aspergillus niger, and
OS Aspergillus awamori.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiaceae; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061, 105351;
[1]
RN
RP SEQUENCE FROM N.A.
RC SPECIES=A.niger;
RX MEDLINE=84261450; PubMed=6204865;
RA Boel E., Hansen M.T., Hjort I., Hoegh I., Pili N.P.;
RT "Two different types of intervening sequences in the glucoamylase gene from Aspergillus niger.";
RL EMOB J. 3:1581-1585(1984).
[2]
RN
RP SEQUENCE FROM N.A.
RC SPECIES=A.niger;
RX MEDLINE=8426105; PubMed=6203744;
RA Boel E., Hjort I., Svensson B., Norris F., Norris K.E., Pili N.P.;
RT "Glucoamylases G1 and G2 from Aspergillus niger are synthesized from two different but closely related mRNAs.";
RL EMOB J. 3:1097-1102(1984).
[3]
RN
RP SEQUENCE OF 25-640, AND COMPARISON OF FORMS G1 AND G2.
RC SPECIES=A.niger;
RX MEDLINE=86136085; PubMed=3081341;
RA Svensson B., Larsen K., Gunnarsson A.;
RT "Characterization of a glucoamylase G2 from Aspergillus niger.";
RL Eur. J. Biochem. 154:497-502(1986).
[4]
RN
RP SEQUENCE OF 25-640.
RC SPECIES=A.niger;
RA Svensson B., Larsen K., Svendsen I., Boel E.;
RT "The complete amino acid sequence of the glycoprotein, glucoamylase G1, from Aspergillus niger.";
RL Carlberg Res. Commun. 48:529-544(1983).
[5]
RN
RP SEQUENCE FROM N.A.
RC SPECIES=A.awamori;
RX MEDLINE=85085934; PubMed=6440004;
RA Nunberg J.H., Meade J.H., Cole G., Lawyer F.C., McCabe P., Schweickart V., Tal R., Wittman V.P., Flatgaard J.E., Innis M.A.;
RT "Molecular cloning and characterization of the glucoamylase gene of Aspergillus awamori.";
RL Mol. Cell. Biol. 4:2306-2315(1984).
[6]
RN
RP REVISIONS.
RC SPECIES=A.awamori;
RA Nunberg J.H., Meade J.H., Cole G., Lawyer F.C., McCabe P., Schweickart V., Tal R., Wittman V.P., Flatgaard J.E., Innis M.A.;
RT Submitted (FEB-1985) to the EMBL/GenBank/DBJ databases.
[7]
RN
RP SEQUENCE OF 1-11 FROM N.A.
RC SPECIES=A.niger; STRAIN=ATCC 10864;
RX MEDLINE=91168302; PubMed=2076554;
RA Fowler T., Berka R.M., Ward M.;
RT "Regulation of the gla gene of Aspergillus niger.";
RL Curr. Genet. 18:537-545(1990).
[8]
RN
RP CONFORMATION OF O-GLYCOSYLATED REGION.
RX MEDLINE=92189576; PubMed=1546955;
RA Williamson G., Belshaw N.J., Williamson M.P.;
RT "O-glycosylation in Aspergillus glucoamylase. Conformation and role in binding.";
RL Biochem. J. 282:423-428(1992).

[9]
RN
RP ACTIVE SITES, AND MUTAGENESIS.
RC SPECIES=A.awamori;
RX MEDLINE=90231978; PubMed=1970434;
RA Sierks M.R., Ford C., Reilly P.J., Svensson B.;
RT "Catalytic mechanism of fungal glucoamylase as defined by mutagenesis of Asp176, Glu179 and Glu180 in the enzyme from Aspergillus awamori.";
RL Protein Eng. 3:193-198(1990).
[10]
RN
RP MUTAGENESIS OF TRP-144.
RC SPECIES=A.awamori;
RX MEDLINE=90046622; PubMed=2510150;
RA Sierks M.R., Ford C., Reilly P.J., Svensson B.;
RT "Site-directed mutagenesis at the active site Trp120 of Aspergillus awamori glucoamylase.";
RL Protein Eng. 2:621-625(1989).
[11]
RN
RP MUTAGENESIS.
RC SPECIES=A.awamori;
RX MEDLINE=93165653; PubMed=8433972;
RA Sierks M.R., Ford C., Reilly P.J., Svensson B.;
RT "Functional roles and substrate locations of Leu177, Trp178 and Asn182 of Aspergillus awamori glucoamylase determined by site-directed mutagenesis.";
RL Protein Eng. 6:75-79(1993).
[12]
RN
RP CHARACTERIZATION OF CATALYTIC DOMAIN.
RX MEDLINE=93277459; PubMed=8503847;
RA Stoffer B., Frandsen T.P., Busk P.K., Schneider P., Svendsen I., Svensson B.;
RT "Production, purification and characterization of the catalytic domain of glucoamylase from Aspergillus niger.";
RL Biochem. J. 292:197-202(1993).
[13]
RN
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 25-495.
RC SPECIES=A.awamori; STRAIN=Var. X100;
RX MEDLINE=92406872; PubMed=1527049;
RA Aleshin A., Golubev A., Firssov L.M., Honzatko R.B.;
RT "Crystal structure of glucoamylase from Aspergillus awamori var. X100 to 2.2-A resolution.";
RL J. Biol. Chem. 267:19291-19298(1992).
[14]
RN
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 25-495.
RC SPECIES=A.awamori; STRAIN=Var. X100;
RX MEDLINE=94253149; PubMed=8195212;
RA Aleshin A., Firssov L.M., Honzatko R.B.;
RT "Refined structure for the complex of acarbose with glucoamylase from Aspergillus awamori var. X100 to 2.4-A resolution.";
RL J. Biol. Chem. 269:15631-15639(1994).
[15]
RN
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 25-495.
RC SPECIES=A.awamori; STRAIN=Var. X100;
RX MEDLINE=94231577; PubMed=8176747;
RA Aleshin A., Hoffman C., Firssov L.M., Honzatko R.B.;
RT "Refined crystal structures of glucoamylase from Aspergillus awamori var. X100.";
RL J. Mol. Biol. 238:575-591(1994).
[16]
RN
RP STRUCTURE BY NMR OF 533-640.
RX MEDLINE=96266494; PubMed=8683599;
RA Sorimachi K., Jacks A.J., Le Gal-Coeffet M.-F., Williamson G., Archer D.B., Williamson M.P.;
RT "Solution structure of the granular starch binding domain of glucoamylase from Aspergillus niger by nuclear magnetic resonance spectroscopy.";
RL J. Mol. Biol. 259:970-987(1996).
[17]
RN
RP STRUCTURE BY NMR OF 533-640.
RX MEDLINE=97341228; PubMed=9195884;
RA Sorimachi K., Le Gal-Coeffet M.-F., Williamson G., Archer D.B., Williamson M.P.;
RT "Solution structure of the granular starch binding domain of

RT Aspergillus niger glucosylase bound to beta-cyclodextrin.";
 RL Structure 5:647-661(1997).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=G1;
 CC IsoId=P04064-1; Sequence=Displayed;
 CC Name=G2;
 CC IsoId=P04064-2; Sequence=VSP_000262;
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X00548; CAA25219.1; -;
 DR EMBL: X00712; CAA25303.1; -;
 DR EMBL: X00712; CAA25304.1; -;
 DR EMBL: X02465; AAB59296.1; -;
 DR EMBL: X02465; AAB59297.1; -;
 DR EMBL: X56442; CAA39825.1; -;
 DR FIR: A29166; A29166.
 DR FIR: A90986; ALASGR.
 DR PDB: 1ACM; 30-SEP-94.
 DR PDB: 1GLM; 31-JUL-94.
 DR PDB: 3GLY; 01-NOV-94.
 DR PDB: 1DOG; 30-APR-94.
 DR PDB: 1ACO; 07-JUL-97.
 DR PDB: 1ACZ; 07-JUL-97.
 DR PDB: 1GAI; 17-AUG-96.
 DR PDB: 1GAI; 17-AUG-96.
 DR PDB: 1KUL; 11-JUL-96.
 DR PDB: 1KUM; 11-JUL-96.
 DR GLYCOSULEDB; P04064; -;
 DR InterPro: IPR002044; CBD_4.
 DR InterPro: IPR008291; Glu-a-glycd_SBD.
 DR InterPro: IPR000165; Glyco_hydro_15.
 DR InterPro: IPR008928; Glyco_trans_6np.
 DR Pfam: PF00686; CBM_20; 1.
 DR Pfam: PF00723; Glyco_hydro_15; 1.
 DR PIRSF: PIRSF001031; Glu-a-glycd_SBD; 1.
 DR PRINTS: PR00736; GLHYDRLASE15.
 DR ProDom: PD001568; CBD_4; 1.
 DR PROSITE: PS00820; GLUCOMYLASe; 1.
 DR HydroLase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KM Alternative splicing; Signal; 3D-structure.
 FT SIGNAL 1 18
 FT PROPEP 19 24
 FT CHAIN 25 640 GLUCOMYLASe G1.
 Query Match 99.0%; Score 2742; DB 1; Length 640;
 Best Local Similarity 99.6%; Pred. No. 9.8e-174;
 Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 181 PLVRNDLSVVAQYWNQGYDLMEEVNGSSFFTLAVGHRALVEGSAFATVAGSSCSWCD5Q 240
 Qy 241 ABEILCYLOSFWTGSTILLANPDSSRSGKANTLLGSIHFDPDEAACDDSTFOCSPRALA 300
 Db 241 ABEILCYLOSFWTGSTILLANPDSSRSGKANTLLGSIHFDPDEAACDDSTFOCSPRALA 300
 Qy 301 NKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPMFLCTAAAEQLYDALYQMD 360
 Db 301 NKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPMFLCTAAAEQLYDALYQMD 360
 Qy 361 KQSLSEVTVSDLPFKALYSDATGYSSSSSTYSIVDAVKTFADPFVSIETHAASNG 420
 Db 361 KQSLSEVTVSDLPFKALYSDATGYSSSSSTYSIVDAVKTFADPFVSIETHAASNG 420
 Qy 421 SMSQYDSDGRLSARDLTWSYALLLTANNRNSVVPASWGTSASSVPGTCAATSAIG 480
 Db 421 SMSQYDSDGRLSARDLTWSYALLLTANNRNSVVPASWGTSASSVPGTCAATSAIG 480
 Qy 481 TYSSVVTSPMSIVATGTTTATPTGSGSVSTSKTTATASKTSTTTS 530
 Db 481 TYSSVVTSPMSIVATGTTTATPTGSGSVSTSKTTATASKTSTTTS 530
 RESULT 2
 AMYG ASPSH STANDARD; PRT; 639 AA.
 AC P22832;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Glucosylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
 DE (1,4-alpha-D-glucan glucosylase).
 GN GLA.
 OS Aspergillus shirousami.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eukaryotes; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxId=5070;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91182400; PubMed=1368603;
 RA Shibuya I., Gomi K., Iimura Y., Takahashi K., Tamura G., Hara S.;
 RT "Molecular cloning of the glucosylase gene of Aspergillus shirousami
 RT and its expression in Aspergillus oryzae.";
 RL Agric. Biol. Chem. 54:1905-1914(1990).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D10460; BAA01254.1; -;
 DR HSSP: P04064; IGAI.
 DR InterPro: IPR002044; CBD_4.
 DR InterPro: IPR008291; Glu-a-glycd_SBD.
 DR InterPro: IPR000165; Glyco_hydro_15.
 DR InterPro: IPR008928; Glyco_trans_6np.
 DR Pfam: PF00686; CBM_20; 1.
 DR Pfam: PF00723; Glyco_hydro_15; 1.
 DR PIRSF: PIRSF001031; Glu-a-glycd_SBD; 1.
 DR PRINTS: PR00736; GLHYDRLASE15.
 DR ProDom: PD001568; CBD_4; 1.
 DR PROSITE: PS00820; GLUCOMYLASe; 1.
 DR HydroLase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KM Signal.
 FT SIGNAL 1 18 POTENTIAL.

Query Match	Best Local Similarity	Score	DB 1	Length	639
Matches	500	Conservative	16	Mismatches	13
				Indels	1
				Gaps	1
FT PROPEP	19	24	BY SIMILARITY.		
FT CHAIN	25	639	GLUCAMYLASE.		
FT BINDING	143	143	SUBSTRATE (BY SIMILARITY).		
FT ACT_SITE	199	199	CATALYTIC BASE (BY SIMILARITY).		
FT ACT_SITE	202	202	GENERAL ACID CATALYST (BY SIMILARITY).		
FT ACT_SITE	203	203	INTERACT WITH SUBSTRATES (BY SIMILARITY).		
FT DISULFID	233	236	BY SIMILARITY.		
FT DISULFID	245	472	BY SIMILARITY.		
FT DISULFID	285	293	BY SIMILARITY.		
FT CARBOHYD	194	194	N-LINKED (GLCNAC. . .) (POTENTIAL).		
FT CARBOHYD	205	205	N-LINKED (GLCNAC. . .) (POTENTIAL).		
FT CARBOHYD	418	418	N-LINKED (GLCNAC. . .) (POTENTIAL).		
SO SEQUENCE	639 AA	68130 MW	E93DAE55EED72326 CRC64		
Query Match	94.6%	Score 2620.5	DB 1	Length 639	
Best Local Similarity	94.3%	Pred. No. 1e-165			
Matches	500	Conservative	16	Mismatches	13
				Indels	1
				Gaps	1
QY	1	MSFSLALSGVYCTGGLANVTSKRTTLDSSLSNEATVARTALINNIGADGAWVSGADGI	60		
DB	1	MSFSLALSGVYCTGGLANVTSKRTTLDSSLSNEATVARTALINNIGADGAWVSGADGI	60		
QY	61	VVASPTNDPQVYFWTWRDGLVLTVDLPENGGTSLSTIENTYISAQAIQVGSINPSG	120		
DB	61	VVASPTNDPQVYFWTWRDGLVLTVDLPENGGTSLSTIENTYISAQAIQVGSINPSG	120		
QY	121	DLSSGAGLGEPEKFNVDETAYTGSWGRPORDGALBATAMIGFQWLLDNGYSTADIYW	180		
DB	121	DLSSG-GLGEPEKFNVDETAYTGSWGRPORDGALBATAMIGFQWLLDNGYSTADIYW	179		
QY	181	PLVRNDLSVYAQYNNQGTGYDLMEEPNGSSFFLIANQHRALVGSAPATAVGSSCWCDSQ	240		
DB	180	PLVRNDLSVYAQYNNQGTGYDLMEEPNGSSFFLIANQHRALVGSAPATAVGSSCWCDSQ	239		
QY	241	APELICYQSFWTGSFFLIANPDSSRSGDANTLLGSIHTFDEAACDSTFQPCSPRALA	300		
DB	240	APOLICYQSFWTGSFFLIANPDSSRSGDNTLLGSIHTFDEAACDSTFQPCSPRALA	299		
QY	301	NHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTLAAREQYDALYQWD	360		
DB	300	NHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTLAAREQYDALYQWD	359		
QY	361	KQGSLEIVTDVSLDFPKALYSDAATGTYSSSSSTYSIYDAVKTFPADGFYSIVETHAASNG	420		
DB	360	KQGSLEIVTDVSLDFPKALYSGAATGTYSSSSSTYSIYSAVTFPADGFYSIVETHAASNG	419		
QY	421	SMSSEYDSDPEQSLARDLMSVYALLLRNRRNSVVPASWGETASSVPGCATSAIG	480		
DB	420	SLSEDFDSDGDELARDLMSVYALLLRNRRNSVVPASWGETASSVPGCAATASAG	479		
QY	481	TYSSVTVSWPSIVATGGTTTATPPTGSGSVYSTSKTTATASKSTTTTSS	530		
DB	480	TYSSVTVSWPSIVATGGTTTATPPTGSGSVYSTSKTTATASKSTTTTSS	529		
RESULT 3					
AMYG_ASPEKA	ID	AMYG_ASPEKA	STANDARD	PRT	639 AA
AC	P23176				
DT	01-NOV-1991	(Rel. 20, Created)			
DT	01-NOV-1991	(Rel. 20, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Glucosylase I precursor (EC 3.2.1.3) (Glucon 1,4-alpha-glucosidase)				
DE	GAI				
GN	Aspergillus kawachi (Aspergillus awamori var. kawachi).				
OS	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
OC	Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.				
OX	NCBI_TaxID=40384;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RT	Hayashida S., Kuroda K., Ohta K., Kuhara S., Fukuda K., Sakaki Y.;				
	"Molecular cloning of the glucosylase I gene of Aspergillus awamori				

RT	var. kaachi for localization of the raw-starch-affinity site." ;	
RL	Agric. Biol. Chem. 53:923-929(1969) .	
RN	[2]	
RP	SEQUENCE OF 494-538	
RA	Hayashida S., Nakahara K., Kuroda K., Miyata T., Iwanaga S. ;	
RT	"Structure of the raw-starch-affinity site on the Aspergillus awamori var. kawachi glucosylase I molecule." ;	
RL	Agric. Biol. Chem. 53:135-141(1969) .	
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.	
CC	-I- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch) .	
CC	-----	
DR	EMBL; D00427; BAA00331.1; -.	
DR	PIR; J00479; J00479.	
DR	GlycoSuiteDB; P23176; -.	
DR	InterPro; IPR002044; CBD_4.	
DR	InterPro; IPR008291; Glu-a-glc5d_SBD.	
DR	InterPro; IPR000165; Glyco_hydro_15.	
DR	InterPro; IPR008928; Glyco_trans_6hp.	
DR	Pfam; PF00686; CBM_20; 1.	
DR	Pfam; PF00723; Glyco_hydro_15; 1.	
DR	PIRSF; PIRSF01031; Glu-a-glc5d_SBD; 1.	
DR	PRINTS; PR00736; GLHYDRLAS15.	
DR	ProDom; PD001568; CBD_4; 1.	
DR	PROSITE; PS00820; GLUCOAMYLASE; 1.	
KW	Hydroxylase; Glycosidase; Polysaccharide degradation; Glycoprotein;	
KW	Signal.	
FT	SIGNAL	1 18
FT	PROPEP	19 24
FT	CHAIN	25 639
FT	DOMAIN	494 538
FT	BINDING	143 143
FT	ACT_SITE	199 199
FT	ACT_SITE	202 202
FT	ACT_SITE	203 203
FT	DISULFID	233 236
FT	DISULFID	245 472
FT	DISULFID	285 293
FT	CARBOHYD	194 194
FT	CARBOHYD	418 418
FT	CARBOHYD	464 464
FT	CARBOHYD	466 466
FT	CARBOHYD	467 467
FT	CARBOHYD	475 475
FT	CARBOHYD	476 476
FT	CARBOHYD	482 482
FT	CARBOHYD	483 483
FT	CARBOHYD	485 485
FT	CARBOHYD	487 487
FT	CARBOHYD	491 491
FT	CARBOHYD	495 495
FT	CARBOHYD	498 498
FT	CARBOHYD	499 499
FT	CARBOHYD	500 500
FT	CARBOHYD	501 501
FT	CARBOHYD	503 503
FT	CARBOHYD	505 505
FT	CARBOHYD	507 507
FT	CARBOHYD	511 511
FT	CARBOHYD	512 512
FT	CARBOHYD	513 513
FT	CARBOHYD	514 514
FT	CARBOHYD	516 516
FT	CARBOHYD	517 517
FT	CARBOHYD	517 517

```

FT CARBOHYD 519 519 O-LINKED (BY SIMILARITY)
FT CARBOHYD 521 521 O-LINKED (BY SIMILARITY)
FT CARBOHYD 523 523 O-LINKED (BY SIMILARITY)
FT CARBOHYD 524 524 O-LINKED (BY SIMILARITY)
FT CARBOHYD 525 525 O-LINKED (BY SIMILARITY)
FT CARBOHYD 526 526 O-LINKED (BY SIMILARITY)
FT CARBOHYD 527 527 O-LINKED (BY SIMILARITY)
FT CARBOHYD 528 528 O-LINKED (BY SIMILARITY)
FT CARBOHYD 529 529 O-LINKED (BY SIMILARITY)
FT CARBOHYD 530 530 O-LINKED (BY SIMILARITY)
FT CARBOHYD 531 531 O-LINKED (BY SIMILARITY)
FT CARBOHYD 533 533 O-LINKED (BY SIMILARITY)
FT CARBOHYD 534 534 O-LINKED (BY SIMILARITY)
SQ SEQUENCE 639 AA; 68271 MW; E112B31A4D8DDB6 CRC64;

Query Match 93.4%; Score 2588.5; DB 1; Length 639;
Best Local Similarity 93.8%; Pred. No. 1,3e-163;
Matches 497; Conservative 15; Mismatches 17; Indels 1; Gaps 1;

QY 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWYSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWYSGADSGI 60

QY 61 VVASPTDNDPYYTWRDGLVLTVDLFRNGDTSLSTIENYISAQAIQGISNPSG 120
DB 61 VVASPTDNDPYYTWRDGLVLTVDLFRNGDTSLSTIENYISAQAIQGISNPSG 120

QY 121 DLSSGAGLGPKNVDEATATGSGWRPQRDGPALRATAMGFQWMLDNGYTSATDIYW 180
DB 121 DLSSGAGLGPKNVDEATATGSGWRPQRDGPALRATAMGFQWMLDNGYTSATDIYW 180

QY 121 DLSSG-GLGEPKFNVDATAATGSGWRPQRDGPALRATAMGFQWMLDNGYTSATDIYW 179
DB 121 DLSSG-GLGEPKFNVDATAATGSGWRPQRDGPALRATAMGFQWMLDNGYTSATDIYW 179

QY 181 PLVRNPLSYAQAQWNOGYDLMEENVNGSSPFT1AVOHRALVEGSAFATVAGSSCSMCDQ 240
DB 180 PLVRNPLSYAQAQWNOGYDLMEENVNGSSPFT1AVOHRALVEGSAFATVAGSSCSMCDQ 239

QY 241 APEILCYLQSGFWTGSFLANFDSSRSKQDNTLLGSIHTEDPPACDSTFQPCSPRALA 300
DB 240 APEILCYLQSGFWTGSFLANFDSSRSKQDNTLLGSIHTEDPPACDSTFQPCSPRALA 299

QY 301 NHEKVDSPRSITLTLDGLSDSEAVAVGRPEDTYTNNGNWFCTLAALAAQYLDALYQMD 360
DB 300 NHEKVDSPRSITLTLDGLSDSEAVAVGRPEDTYTNNGNWFCTLAALAAQYLDALYQMD 359

QY 361 KQGSLEVTDVSLDFPFLALYSDATGTYSSTSSSTYSIVDAVKTGFADGFVIVETHAASNG 420
DB 360 KQGSLEVTDVSLDFPFLALYSDATGTYSSTSSSTYSIVDAVKTGFADGFVIVETHAASNG 419

QY 421 SMSBQYDKSDGQSLARDLTWYSYALLLTANNRRNSVVPASGRTSASVPGTCAATSAIG 480
DB 420 SMSBQYDKSDGQSLARDLTWYSYALLLTANNRRNSVVPASGRTSASVPGTCAATSAIG 479

QY 481 TYSSVTVTWSPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTS 530
DB 480 TYSSVTVTWSPSIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTS 529

RESULT 4
AMYG_ASPPOR STANDARD; PRT; 612 AA.
ID AMYG_ASPPOR STANDARD; PRT; 612 AA.
AC P36914;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GLAA.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillius.
OK NCBI_TaxID=5062;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91254744; PubMed=1368680;

```

```

RA Hata Y., Kitamoto K., Gomi K., Kumagai C., Tamura G., Hara S.;
RT "The glucoamylase cDNA from Aspergillus oryzae: its cloning,
RT nucleotide sequence, and expression in Saccharomyces cerevisiae.";
RL Agric. Biol. Chem. 55:941-949(1991).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=92104497; PubMed=1761224;
RA Hata Y., Tsuchiya K., Kitamoto K., Gomi K., Kumagai C., Tamura G.,
RA Hara S.;
RT "Nucleotide sequence and expression of the glucoamylase-encoding gene
RT (glia) from Aspergillus oryzae.";
RL Gene 108:145-150(1991).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=RIB 40;
RA Hara S., Tamura G., Kumagai C., Gomi K., Kitamoto K., Tsuchiya K.,
RA Hata Y.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D01035; BAA00841.1; -
DR EMBL; D10598; BAA01540.1; -
DR PIR; J01346; J01346.
DR HSSP; P04064; IGAI.
DR InterPro; IPR002044; CBD_4.
DR InterPro; IPR008291; Glu-a-glycd_SBD.
DR InterPro; IPR000165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_crtans_6hp.
DR Pfam; PF00686; CBM_20; 1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PIRSF; PIRSF01031; Glu-a-glycd_SBD; 1.
DR PRINTS; PRO0736; GLHYDRASE15.
DR PRODOM; PD001568; CBD_4; 1.
DR PROSITE; PS00820; GLUCOAMYLASE; 1.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT PROPEP 20 25 BY SIMILARITY.
FT CHAIN 26 612 GLUCOAMYLASE.
FT BINDING 146 146 SUBSTRATE (BY SIMILARITY).
FT ACT_SITE 202 202 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 205 205 GENERAL ACID CATALYST (BY SIMILARITY).
FT ACT_SITE 206 206 INTERACT WITH SUBSTRATES (BY SIMILARITY).
FT DISULFID 236 239 BY SIMILARITY.
FT DISULFID 248 475 BY SIMILARITY.
FT DISULFID 288 296 BY SIMILARITY.
FT CARBOHYD 39 39 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 612 AA; 65486 MW; CD7B23B5FA978F97 CRC64;

Query Match 69.8%; Score 1935; DB 1; Length 612;
Best Local Similarity 70.3%; Pred. No. 1,9e-120;
Matches 367; Conservative 59; Mismatches 90; Indels 6; Gaps 4;

QY 1 MSFRSLALSGVCTGLA--NVISKRATLDSWLSNEATVARTAILNNIGADGAWYSGAD 58
DB 2 VFSFSLRALALGSSVLAQVPLRQATGADTLSTEANFSROAILNNIGADGQSAQASAP 61

QY 59 GIVVASPTDNDPYYTWRDGLVLTVDLFRNGDTSLSTIENYISAQAIQGISNPS 118
DB 62 GVIVASPSKSDPDYFTWTRDGLVLTVDLFRNGDADLPIIEFISQARIQGISNP 121

QY 119 SGDLSSGAGLGPKNVDEATATGSGWRPQRDGPALRATAMGFQWMLDNGYTSATDI 178

```

```
Db 122 SGALSSS-GAGEPFVNDFTAWGRPDRDGLATAMISFGEWLVENSHTSIATDL 180
Qy 179 VMPVLRNDLSYVAQYQVNTQGYDLMEEVNCGSPFTIAVOHRLVGSFAFATVAGSSCMCD 238
181 VMPVLRNDLSYVAQYQVNTQGYDLMEEVNCGSPFTIAVOHRLVGSFAFATVAGSSCMCD 240
Qy 239 SOAPEILICYQSFWTGSFILLANPDSRSGKDANTLLGSIHTFDEPAACDSTFQPCSPRA 298
241 SOAPEILICYQSFWTGSFILLANPDSRSGKDANTLLGSIHTFDEPAACDSTFQPCSPRA 300
Qy 299 LANHKEVVDSPRSYLTINDGSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOYLALYQ 358
301 LANHKEVVDSPRSYLTINDGSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOYLALYQ 360
Qy 359 WDKQGSLEVTVDVSLDFPKALYSDAATGTVSSSSSTYSIVDAVKTFADEFSIVETHAAS 418
361 WDKQGSLEVTVDVSLDFPKALYSDAATGTVSSSSSTYSIVDAVKTFADEFSIVETHAAS 420
Db 419 NGSMSEQYDKSDGEQLSARDLTWMSYALLTANNRRNSVVPASWGETSASSVPGTCAATSA 478
421 TGSMAEQYTKTDGSGQTSARDLTWMSYALLTANNRRNSVVPASWGETSASSVPGTCAATSA 480
Qy 479 IGTYSVTVTSWPSIVATGTF--TTATPTGSSGVSSTSKTT 518
481 IGTYSVTVTSWPSIVATGTF--TTATPTGSSGVSSTSKTT 521
```

RESULT 5

AMYG_NEUCR STANDARD; PRT; 626 AA.

```
AC P14804; Q9P5U5;
DT 01-APR-1990 (Rel. 14, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glucosylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GLA-1 OR B5022.70.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxId=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A / FGSC 987;
RX MEDLINE=94037144; PubMed=8221928;
RA Stone P.J., Makoff A.U., Parish J.H., Radford A.;
RT "Cloning and sequence analysis of the glucanase gene of Neurospora
RT crassa."
RL Curr. Genet. 24:205-211 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A / FGSC 987;
RX MEDLINE=22542210; PubMed=12655011;
RA Mannhaupt G., Montreone C., Haase D., Mewes H.-W., Aign V.,
RA Hoesel J.D., Fartmann B., Nyakatura G., Kempen F., Maier J.,
RA Schulte U.;
RT "What's in the genome of a filamentous fungus? Analysis of the
RT Neurospora genome sequence."
RL Nucleic Acids Res. 31:1944-1954 (2003).
RN [3]
RP SEQUENCE OF 36-65.
RC STRAIN=74-OR23-1A / FGSC 987;
RA Koh-Luar S.I., Parish J.H., Bleasby A.U., Pappin D.J.C., Atanley K.,
RA Johansen F.E., Radford A.;
RT "Exported proteins of Neurospora crassa: 1-glucosylase."
RL Enzyme Microb. Technol. 11:692-695 (1989).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>
CC or send an email to license@ebi.ac.uk).

```
DR EMBL; X67291; CAA47707.1; -.
DR EMBL; AL355932; CAB91426.1; -.
DR HSSP; P04064; 1GAI.
DR InterPro; IPR002044; CBD_4.
DR InterPro; IPR008291; Glu-4-glycosid SHD.
DR InterPro; IPR000165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_trans_6mp.
DR Pfam; PF00686; CBM_20; 1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PIRSF; PIRSF01031; Glu-4-glycosid_SBD; 1.
DR PRINTS; PR00735; GLHYDRASE15.
DR ProDom; PD001568; CBD_4; 1.
DR PROSITE; PS00820; GLUCOMYLASE; 1.
KM Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KM Signal.
FT SIGNAL. 1 19
FT PROPEP 20 35
FT CHAIN 36 626
FT BINDING 155 155
FT ACT SITE 211 211
FT ACT_SITE 214 214
FT ACT_SITE 215 215
FT CARBOHYD 106 106
FT CARBOHYD 206 206
FT CARBOHYD 217 217
FT CONFLICT 82 82
FT CONFLICT 550 550
FT CONFLICT 560 560
SQ SEQUENCE 626 AA; 66474 MW; 54E5BDBA7A3E349 CRC64;
```

Query Match 56.9%; Score 1577; DB 1; Length 626;
Best Local Similarity 57.6%; Pred. No. 8, 6e-97;
Matches 306; Conservative 85; Mismatches 130; Indels 10; Gaps 6;

```
Qy 2 SFRSLALSLGLVGTGLANVSKRATLDSMNSNATVARTALINIGADGAMVSGSIV 61
Db AFOAVLGLPDLPEKRRHSDIKR-SVDSYIQTTPTAOKNLCNIGASGRASGASGV 71
Qy 62 VASPSSTNDPQFYTWTRDSGLVLTVDLFRNG-DTSLSTTENYISAQIVGSIENPSG 120
Db VASPSKSSPDYWTWTRDALVTKLVDEFNTDNTNTQNTQAVAAQKIQGVENPSG 131
Qy 121 DLSSGAGLGEPRKNNVDETAVTSGWGRPQRDGPALRTAMIGCOMLLDNGYSTATDIW 180
Db SLNCGAGLGEPRKNNVDETAVTSGWGRPQRDGPALRTAMIGCOMLLDNGYSTATDIW 191
Qy 181 EIVRNRLSYVAQVQVNTQGYDLMEEVNCGSPFTIAVOHRLVGSFAFATVAGSSCMCD 240
Db EIVRNRLSYVAQVQVNTQGYDLMEEVNCGSPFTIAVOHRLVGSFAFATVAGSSCMCD 251
Qy 241 APRILICYQSFWTGS-FILLANPDSRSGKDANTLLGSIHTFDEPAACDSTFQPCSPRA 299
Db APRILICYQSFWTGS-FILLANPDSRSGKDANTLLGSIHTFDEPAACDSTFQPCSPRA 311
Qy 300 ANHKEVVDSPRSYLTINDGSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOYLALYQ 359
Db ANHKEVVDSPRSYLTINDGSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOYLALYQ 370
Qy 360 DKGQGSLEVTVDVSLDFPKALYSDAATGTVSSSSSTYSIVDAVKTFADEFSIVETHAAS 419
Db DKGQGSLEVTVDVSLDFPKALYSDAATGTVSSSSSTYSIVDAVKTFADEFSIVETHAAS 430
Qy 420 GNSMEQYDKSDGEQLSARDLTWMSYALLTANNRRNSVVPASWGETSASSVPGTCAATSA 479
Db GNSMEQYDKSDGEQLSARDLTWMSYALLTANNRRNSVVPASWGETSASSVPGTCAATSA 490
```


QY 480 GYSSVTVTSMPSIVATGTTT-ATPTSGS-----VTSSTKTATASKT 524
 DB 491 GSATATATATSPANLTPASTTTPPTGTGCAADHEVLVTNEKVTTSYGT 541

RESULT 6
 ID AMYG HORRE STANDARD; PRT; 616 AA.
 AC 003045;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE Glucoamylase P precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
 DE (1,4-alpha-D-glucan glucohydrolase).
 GN GAMP.
 OS Hormoonis resiniae (Creosote fungus) (Amorphotheca resiniae).
 OC Eukaryota; Fungi; Ascomycota; Ascomycota incertae sedis;
 OC Amorphothecaceae; Amorphotheca.
 OX NCBI_TaxId=5101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 20495;
 RX MEDLINE=93146382; PubMed=1490604;
 RA Joutejoki V.V., Torkkeli T.K.;
 RT "Glucoamylase P gene of Hormoonis resiniae: molecular cloning,
 RT sequencing and introduction into Trichoderma reesei";
 RL FEMS Microbiol. Lett. 78:237-243 (1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 20495;
 RX MEDLINE=93165035; PubMed=8358830;
 RA Vainio A.E.I., Torkkeli H.T., Tuusa T., Aho S.A., Fagerstrom B.R.,
 RA Korhola M.P.;
 RT "Cloning and expression of Hormoonis resiniae glucoamylase P cDNA in
 RT Saccharomyces cerevisiae";
 RL Curr. Genet. 24:38-44 (1993).
 RN [3]
 RP SEQUENCE OF 72-76, AND CHARACTERIZATION.
 RX MEDLINE=90338987; PubMed=2116499;
 RA Fagerstrom R., Vainio A.E.I., Suoranta K., Pakula T., Kalkkinen N.,
 RA Torkkeli H.T.;
 RT "Comparison of two glucoamylases from Hormoonis resiniae";
 RL J. Gen. Microbiol. 136:913-920 (1990).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

EMBL; X68143; CAA48243.1; -
 EMBL; X67708; CAA47945.1; -
 DR PIR; S33908; S33908.
 DR HSSP; P04064; 1GAI.
 DR InterPro; IPR002044; CBD_4.
 DR InterPro; IPR008291; Glu-a-glcSD_SBD.
 DR InterPro; IPR000165; Glyco_hydro_15.
 DR InterPro; IPR008928; Glyco_trans_6np.
 DR Pfam; PF00686; CBM_20; 1.
 DR Pfam; PF00723; Glyco_hydro_15; 1.
 DR PIRSF; PIRSF001031; Glu-a-glcSD_SBD; 1.
 DR PRINTS; PR00736; GHYDRLASES15.
 DR ProDom; PD001568; CBD_4; 1.
 DR PROSITE; PS00820; GLUCOAMYLASE; FALSE NEG.
 KW Hydroxylase; Glycosidase; Polyaccharide degradation; Glycoprotein;
 KW Signal.

FT SIGNAL 1 29
 FT CHAIN: 30 616
 FT BINDING 149 149
 FT ACT SITE 205 205
 FT ACT_SITE 208 208
 FT ACT_SITE 209 209
 FT CARBOHYD 200 200
 FT CARBOHYD 427 427
 SQ SEQUENCE 616 AA; 66432 MW; B5FADCEBBB152FB CRC64;

Query Match 47.7%; Score 1321; DB 1; Length 616;
 Best Local Similarity 48.6%; Pred. No. 6, 8e-80;
 Matches 257; Conservative 88; Mismatches 150; Indels 34; Gaps 6;

QY 14 CIGLANVIS-----KRALDLSLSEATVARTALINNIGADGAWGSDGIVVA 63
 DB 8 CGAUSLCSGLIAAPTELKARDLSFISERAILAQGLANNIGPDSAVPGAGFVVA 67
 QY 64 SPSTNDPDYFYTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYVGISNPSGD-L 122
 DB 68 SPKSNAPDYFYTWSRDSALTLKMIIDEFLGNTTLOTIIEOYIHAQAVIQTVSNPGTFL 127
 QY 123 SSAGLGEPEKFNVDETAYGSGWRPQDGPALRATMIGFGOMLDNGYTTATDIWPL 182
 DB 128 PGVGIGEBKFWVDGTRFNGPWRPQDGPALRALMTYSWMLIRNGOFABEKTIWPI 187
 QY 183 VNRDLSVYVQVNYNGTGYDLMEEVNCSFFETIAVQHRALVGSAPATAVSSGSCSOAP 242
 DB 188 IANDLSYVQVNYNGSGFDLMEETVYASSPFTIONQHRALVGSAGLADLGTCTGCD-QAP 246
 QY 243 EILCYLQSFMTGSGFILANF--DSRSKQDANTLLGSIHTFDEPAACDSTFQPCSPRAL 300
 DB 247 EYLCFLQSFMTGNGXIVSNNVANNRGRTDGNSTLGIASFIDIDAYCDSPTLQCHSGSLA 306
 QY 301 NHKEVVDSPRSYITLNDGSDSEAVAGRYPEDTYNGNPWFCLTAAABQLYDALYQMD 360
 DB 307 NFKVLDTDFRNLTYTINAGIPEGQVAVGRVYAEVYVGNPNWYLITAAAEFLYDAVAQWK 366
 QY 361 KQGSLEVTDVSLDFEALSDAATGYSS--SSSTYSIIVDAVKTFADPVSIVETHAAS 418
 DB 367 ARHVLTVDETSIAFFDIDYPEVTREYKSGNNSPPAQIMDAVTAVADSVYALAEKYPIS 426
 QY 419 NGSMSEQYKSDQEQSLARDLFTMSYALTLTANNRRNSVPASGETSASSVPGTCATSA 478
 DB 427 NGSLSQPRKRDTPSLAIDLFTMSYAFITMOKRAGQYPPSSGSGNALLPPTTCASST 486
 QY 479 IGTYSSTVTVTSMPSIVATGTTTATPTGSGVTSSTKTATASKTSTT 527
 DB 487 PGYI-----TPATPAGAPNVTSSQVSIFFINATT 517

RESULT 7
 ID AMYG SCHPO STANDARD; PRT; 450 AA.
 AC 060087; 096WS5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Probable glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
 DE glucosidase) (1,4-alpha-D-glucan glucohydrolase) (melotic expression
 DE upregulated protein 17).
 GN MEU17 OR SPBC148.05C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxId=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N.T., Hayes S., Baker S., Baahman D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volkart G., Aert R., Roben J., Gymnopoulos B., Melitsis I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M., Beyer P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucac M., Rocher M., Gallardin C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Carrutli L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.; RT "The genome sequence of Schizosaccharomyces pombe."; RL Nature 415:871-880(2002).

RP SEQUENCE OF 210-450 FROM N.A.

RC STRAIN=CD16-1;

RA MEDLINE=21270454; PubMed=11376151;

RA Matanabe T., Miyashita K., Saito T.T., Yoneki T., Kakihara Y., Nabeshima K., Kishi Y.A., Shimoda C., Nojima H.;

RT "Comprehensive isolation of meiosis-specific genes identifies novel proteins and unusual non-coding transcripts in Schizosaccharomyces pombe.";

RL Nucleic Acids Res. 29:2327-2337(2001).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.

CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AL022305; CA18423.1; -

DR EMBL: AB054301; BAB60870.1; -

DR F1R: T39433; T39433.

DR HSP: P04064; IGAI.

DR GeneDB: SPBC14C8.05c; -

DR InterPro: IPR000165; Glyco_hydro_15.

DR InterPro: IPR008928; Glyco_trans_6hp.

DR Pfam: PF00723; Glyco_hydro_15; 1.

DR PRINTS: PR00736; GLYHDLASE15.

KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein; KM Signal; Meiosis.

FT SIGNAL 1 16 POTENTIAL.

FT PROPEP 17 28 BY SIMILARITY.

FT CHAIN 29 450 PROBABLE GLUCOAMYLASE.

FT BINDING 147 147 SUBSTRATE (BY SIMILARITY).

FT ACT_SITE 203 203 CATALYTIC BASE (BY SIMILARITY).

FT ACT_SITE 206 206 GENERAL ACID CATALYST (BY SIMILARITY).

FT ACT_SITE 207 207 INTERACT WITH SUBSTRATES (BY SIMILARITY).

FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 450 AA; 51163 MW; 31C5BFA28BA785FE CRC64;

Query Match 28.2%; Score 781.5; DB 1; Length 450; Best Local Similarity 37.6%; Pred. No. 1.9e-44; Matches 179; Conservative 68; Mismatches 166; Indels 63; Gaps 12;

QY 7 LAISGLVCTG---LANVISKRATLDSWLSNENATVARTAILNNIGADGANVSGADSGIVVA 63
Db 8 LILGGVSAESLSLPKRKEKASMDWTQOQKIAIGHNIIJGSGMAKADINPCIIA 67
QY 64 SPSTNDPQFYWTRPSGLVLTVDLPFRNGPTSLSTENIYSAQAIYOGISNPSGDS 123
Db 68 SPSTDPDYVQWVRSAIITMTILDFREGDKLEPIIVKYNDEWRLQKYPNPGDPR 127
QY 124 SGAGLGEPPFNVDETAYTSGWGRPQRDGPALRATAMIGCGWMLDNGYSTATDIWPLV 183
Db 128 AG-GLEPPFNVDGTYDDGWRQPDSDPALRAIATKMYNLFENGKEVHEVTIEMAV 186
QY 184 RNDLSVAQYMNQGTVDLBEVNGSSFTIAYOHRVLVGSAPATLVGSSCSMCSDAPB 243
Db 187 LADLDYTAHMTWASFDLBEIEDVYFTLAQVKRAMOGTAFKIG-----APD 237
QY 244 -----ILCYQSFV-TGSPFLANFDS--SSGQDANTLGSHT--FPDEACD 287
Db 238 QAAALYRTTEPIDLKGFEWDPQMGVIKGYKRVDSGLDCSTLLSLVSNFED----- 291
QY 288 DSTFQPCSPRALNHRKVVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYN-----GNPWF 342
Db 292 -----WHILPTLLKLGETMTRDYPVNQGWQ-----AMGRPEVDYGVSKSGNPWF 339
QY 343 LCTLAABEQYDALYQMDKQGLSEVTDVSLDPFK--ALYSDAATGTYSSSSSTYSIVA 400
Db 340 ICTSSAETIYKAIAYYDKGLPELTENVIHFFMKFAEFGD-----PYWVSYIRKN 390
QY 401 VKTFADGFSIYETHAASNGSMSEYDKSDGEOLSRDLTWYSYALLTANRNSV 456
Db 391 MHTYADNFKAAVAFQHPNGMSSEQPSRDDGQKGRDLTWYSYSLNATYRRAI 446

RESULT 8

AMYG_RHIOR

ID AMYG_RHIOR STANDARD; PRT; 604 AA.

AC P07683;

DT 01-APR-1988 (Rel. 07, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Glucoamylase precursor (BC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)

DE (1,4-alpha-D-glucan glucohydrolase).

OS Rhizopus oryzae (Rhizopus delamar).

OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;

OC Rhizopus.

OX NCBI_TaxID=64495;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SAM0034;

RA Ashikari T., Nakamura N., Tanaka Y., Kiuchi N., Shibano Y.,

RA Tanaka T., Amachi T., Yoshizumi H.;

RT "Rhizopus raw-starch-degrading glucoamylase: its cloning and

RT expression in yeast".

RL Agric. Biol. Chem. 50:957-964(1986).

RN [2]

RP HOMOLOGY, AND PREDICTED SECONDARY STRUCTURE.

RA Tanaka Y., Ashikari T., Nakamura N., Kiuchi N., Shibano Y.,

RA Amachi T., Yoshizumi H.;

RT "Comparison of amino acid sequences of three glucoamylases and their

RT structure-function relationships".

RL Agric. Biol. Chem. 50:965-969(1986).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-

CC glucose residues successively from non-reducing ends of the chains

CC with release of beta-D-glucose.

CC -1- MISCELLANEOUS: Rhizopus glucoamylase exists in multiple forms,

CC Gluc 1, Gluc 2, and Gluc 3, all of which hydrolyze gelatinized

CC starch at similar rates, but only the largest one (Gluc 1) is able

CC to adsorb raw starch.

CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL, D000049; BAA00033.1; -

DR HSP; JP0001; JP0001.

DR HSP; P04064; 1GAI.

DR InterPro; IPR005036; CBM_21.

DR InterPro; IPR000165; Glyco_hydro_15.

DR InterPro; IPR008928; Glyco_trans_6np.

DR Pfam; PF03370; CBM_21; 1.

DR Pfam; PF00723; Glyco_hydro_15; 1.

DR PRINTS; PR00736; GUHYDRASE15.

DR PROSITE; PS00820; GLUCOMYLAASE; 1.

KW Hydroxylase; Glycosidase; Polysaccharide degradation; Glycoprotein; Signal.

FT SIGNAL 1 25

FT CHAIN 26 604

FT CHAIN 159 604

FT CHAIN 116 604

FT DOMAIN 26 109

FT BINDING 116 604

FT ACT_SITE 279 279

FT ACT_SITE 336 336

FT ACT_SITE 339 339

FT ACT_SITE 340 340

FT ACT_SITE 340 340

FT ACT_SITE 604 AA; 65162 MW; 78421F1AA03ADB9 CRC64;

Query Match 26.2%; Score 725; DB 1; Length 604; Best Local Similarity 36.8%; Pred. No. 1.6e-40; Matches 165; Conservative 71; Mismatches 170; Indels 42; Gaps 10;

QY 25 AILDMSLNENATVARTAILNIGADGAWSGADSGIVASPSIDNDPFFYTRDSCGLV 84

DB 168 STISSWIKQEGISRFAMLRINP-----PGSATGFIAASLSTAGPDYVYATRDALTS 222

QY 85 KTLVDLFR--NDDTSLSTIENYISAOAVOGISNNSGDLSSGAGLGEKPNVDEAYT 141

DB 223 NVIVYEINTTLTSGNKTILNVLKDYVPSVKTOSTYVCN-----CLGEKPNVDASGYT 276

QY 142 GSWGRPDQDPALPAPATAMIGFG--QMLDNGYSTATDIWPLVRNDLSVYAQVWNTGYD 200

DB 277 GAWGRQNDGPAPARATFTIFADSYLTQTDASVYTGTLKPAIFKLDYVWVWNSNCFD 336

QY 201 LMEEVNGSFFTTAVQHRALVEGSAFATVAGSS--CSWCDSQAPETLCYQSGFWTGSFIL 258

DB 337 LMEEVNGVHFYTLTVMWKKGLLGADFAKRGDSTRASTYSTASTIANKISSFWSSNMW 396

QY 259 ANPDS-----SRSGDANTL---LGSITPDBAACDSTQPCGPRALAHKEVYDS 308

DB 397 IQVSQSVTGVSCKKGLDVSTFLANLGSV-----DDGFTPGSEKILATAVAVEDS 447

QY 309 FRGIYTLNDGLSDESAVAVRYPEDTY-----YNGNPMFLCTLAABQLDALYOMDKOG 363

DB 448 FALPLPINKRLPSYLGNSIGRYEDDTYNGNGNSGNSWFLAVGYALRYALKENWGNG 507

QY 364 SLEVTVDLDFKALYSDAATG--TYSSSSSTYSIVDAVTFADGFVSIIVETHAASNGSM 422

DB 508 GATVSSISLPEFKKFDSSATSGKKYTGTSDFNNLAQNALADRFLSTYQLAHANNHNSL 567

QY 423 SEQYDASDGEQLARDLTWSYALTLTAN 450

DB 568 AAEFRTTGLSTGARDLTWASHASLTAS 595

RESULT 9

AMTH_SACFI

ID AMTH_SACFI STANDARD; PRT; 519 AA.

AC P26989; P78745;

DT 01-AUG-1992 (Rel. 23, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Glucanase GLA1 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucanhydrolase).

GN GLA1

OS Saccharomycopsis fibuligera (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycopsidaceae; Saccharomycopsis.

OX NCBI_Taxid=4944;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KZ;

RX MEDLINE=92137640; PubMed=1840532;

RA Hostinova E.; Balanova J.; Gasperik J.;

RT "The nucleotide sequence of the glucanase gene GLA1 from RT Saccharomycopsis fibuligera KZ."

RL PEMS Microbiol. Lett. 67:103-108(1991).

RN [2]

RP REVISIONS.

RC STRAIN=KZ;

RA Hostinova E.;

RL Submitted (Feb-1997) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.

CC -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES. ONLY DIFFERS IN 11 POSITIONS WITH GLUCOMYLAASE GLU1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X58117; CAA4120.1; -

DR HSP; P08017; IAYX.

DR InterPro; IPR000165; Glyco_hydro_15.

DR InterPro; IPR008928; Glyco_trans_6np.

DR Pfam; PF00723; Glyco_hydro_15; 1.

DR PRINTS; PR00736; GUHYDRASE15.

DR PROSITE; PS00820; GLUCOMYLAASE; 1.

KW Hydroxylase; Glycosidase; Polysaccharide degradation; Glycoprotein; Signal.

FT SIGNAL 1 27

FT CHAIN 28 519

FT BINDING 166 166

FT ACT_SITE 234 234

FT ACT_SITE 237 237

FT ACT_SITE 238 238

FT CARBOHYD 115 115

FT CARBOHYD 127 127

FT CARBOHYD 205 205

FT CARBOHYD 205 205

FT CARBOHYD 519 AA; 57542 MW; A15A009A7640053C CRC64;

Query Match 24.2%; Score 671.5; DB 1; Length 519; Best Local Similarity 35.2%; Pred. No. 4.3e-37; Matches 173; Conservative 73; Mismatches 169; Indels 77; Gaps 16;

QY 24 RATLDMSLNENATVARTAILNIGADGAWSGADSGIVASPSIDNDPFFYTRDSCGL- 82

DB 42 RTDLERFLDKQKQVSLYLLQNLAYPEQFNDSVPETVLASPETSNDPYQYQTRDSALT 101

QY 83 ---VLTVDLFRNGDTSLSLTENYISAOAVOGISNNSG--DLSSGAGLGEKPNVDE 137

DB 102 FLTVLSELSD--NNFTTLAKAVEYIYNISYVLTQRTSNPSGSDDNHNGLGEKPNVDE 159

QY 138 TATYSGWRPDQDPALPAPATAM-----IGFGQML-----DNGYSTATDIWPLVR 184

DB 160 SATYGAWRPDQDPALPAPATAM-----IGFGQML-----DNGYSTATDIWPLVR 218

QY 185 NDLSSVYAQVWNTGYDLMEEVNGSFFTTAVQHRALV-----EGSAFATVAGSSCG 235

DB 219 PDLEVIYGVWDSITGFDLMEENQGRHFTSLVQOKALAVANDIAKSPDDGDFANTLSSTAS 278

RA Shima H., Inui M., Akada R., Yamashita I.;
 RT "upstream regions of the yeast glucosylase gene which are required
 for efficient transcription.";
 RL Agric. Biol. Chem. 53:749-755 (1989).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 glucose residues successively from non-reducing ends of the chains
 with release of beta-D-glucose.
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X02649; CAA26487.1; ALT_INIT.
 DR EMBL: D00428; BAA00332.1; -.
 DR PIR: A21896; ALBYG.
 DR HSSP: P08017; IAYX.
 DR InterPro: IPR000165; Glyco_hydro_15.
 DR InterPro: IPR008928; Glyco_trans_6np.
 DR Pfam: PF00723; Glyco_hydro_15; 1.
 DR PRINTS: PR00736; GLHYDRASE15.
 DR PROSITE: PS00820; GLUCOMYLAZE; 1.
 KM Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW Signal; Multigene family.
 FT SIGNAL 1 21
 FT CHAIN 22 767
 FT DOMAIN 22 347
 FT DOMAIN 348 691
 FT BINDING 692 767
 FT ACT_SITE 518 518
 FT ACT_SITE 521 521
 FT ACT_SITE 522 522
 FT CARBOHYD 308 308
 FT CARBOHYD 322 322
 FT CARBOHYD 414 414
 FT CARBOHYD 423 423
 FT CARBOHYD 434 434
 FT CARBOHYD 513 513
 FT CARBOHYD 546 546
 FT CARBOHYD 645 645
 FT CARBOHYD 650 650
 FT CARBOHYD 720 720
 FT CARBOHYD 741 741
 SQ SEQUENCE 767 AA; 82488 MW; A5P29E2427EDB593 CRC64;
 Query Match 18.8%; Score 522; DB 1; Length 767;
 Best Local Similarity 30.5%; Pred. No. 5,5e-27;
 Matches 139; Conservative 80; Mismatches 177; Indels 60; Gaps 15;
 QY 1 MSFRSLALSGVCTGLANVTSKATLDSWLSMEATVAFARAILNNGADGAWTSGDSGI 60
 DB 313 VQLRDVLAMGTYV-D-SNCAMDSALAEMLQKOKVIRIRIENGPSAVIPS-ILPGV 370
 QY 61 VVASPSTNDNDFYRTWTRDSGLVLTVDLFRNGDTSLSTINENYSAQIVOGISNPSG 120
 DB 371 VVASPQTHDHYFYQWIRDSALITNSIVS--HSADPA-ITTLQYLVNPHLRITNN--- 424
 QY 121 DLSSGAG-----LGEKRVDEIATVTSWGRPQRDGLPRLATAMIGFQWILNDG--- 170
 DB 425 TLGAGGYTNDVALGDPKKNVNTAFTEBPGRPQNDGPALRSIALIKIIDYIKOSTDL 484
 QY 171 -----TSTATDVLVPLVRNDLSVAQYMNQGTVDLMEEVNNGSSFTTAVOHALVSGSA 225
 DB 485 GAKYPPQSTH-DLFPDILVRWDLRFIIIDHNNSSGDLMEEVNNGHFFTLVQLSAVDRSL 543
 QY 226 FATAVSSCSGWC-----SQAPEILCYL-----QSFWTGSFIILANF-----DSSRSGKDA 270

DB 544 YFNASERSPPFVBEILQTRRDISKFLVPDANGFINKY---NYIVETPMIADTLRSGLDI 600
 QY 271 NTLGSIHTFDEPAACDDSTFQPCSPRALANKKEVDSFRSITYTLNDGSDSEAVAVGRY 330
 DB 601 STLLANTYHADPSA-SHPFDINDPAVANTLJHMLHNRISITYPINDSKNTGIALGRY 659
 QY 331 PEDTY-----YNGNPWFCTLAAEOLYALYOMDKQSLSEVTVDLDFPKALYSDAATG 385
 DB 660 PEDVYDYGVGSGNPVATLTCAASTLYQLYRHISEQHDVLVPMNNDGSMNMFSELVPS 719
 QY 386 TYSS-----SSSTYSIVDAVTFPFDGAV 409
 DB 720 NTLTGNDGEGYLLEFNTPAFQTLQIKIQLADSF 755
 RESULT 14
 ID AMY1_SACDI STANDARD; PRT; 768 AA.
 AC P29760;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glucosylase 62 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
 DE (1,4-alpha-D-glucan glucosylase) (Gall).
 GN SYA2 OR DEX1.
 OS Saccharomyces diastaticus (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NC NCB1_Taxid=41870;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91276266; PubMed=8017901;
 RA Kim K., Bajszar G., Lee S.Y., Knudsen F., Mattoon J.R.;
 RT "Cloning of a new allelic variant of a Saccharomyces diastaticus
 glucosylase gene and its introduction into industrial yeasts.";
 RL Appl. Biochem. Biotechnol. 44:161-185(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91276266; PubMed=2055484;
 RA Lambrechts M.G., Pretorius I.S., Solliet P., Marmur J.;
 RT "Primary structure and regulation of a glucosylase-encoding gene
 (STA2) in Saccharomyces diastaticus.";
 RL Gene 100:95-103(1991).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 glucose residues successively from non-reducing ends of the chains
 with release of beta-D-glucose.
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M60650; AAA35107.1; -.
 CC EMBL: M90490; AAA20560.1; -.
 CC PIR: J00474; J00474.
 CC HSSP: P08017; IAYX.
 CC InterPro: IPR000165; Glyco_hydro_15.
 CC InterPro: IPR008928; Glyco_trans_6np.
 CC Pfam: PF00723; Glyco_hydro_15; 1.
 CC PRINTS: PR00736; GLHYDRASE15.
 CC PROSITE: PS00820; GLUCOMYLAZE; 1.
 KM Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW Signal; Multigene family.
 FT SIGNAL 1 21
 FT CHAIN 22 768
 FT DOMAIN 22 348
 FT DOMAIN 349 692
 FT BINDING 693 768
 FT BINDING 456 456
 SUBSTRATE (BY SIMILARITY).

FT ACT SITE 519 519 CATALYTIC BASE (BY SIMILARITY).
FT ACT SITE 522 522 GENERAL ACID CATALYST (BY SIMILARITY).
FT ACT SITE 523 523 INTERACT WITH SUBSTRATES (BY SIMILARITY).
FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 514 514 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 164 164 MISSING (IN REF. 1).
FT CONFLICT 624 624 D -> N (IN REF. 1).
SQ SEQUENCE 768 AA; 82586 MW; 3PAC172C128A0C6F CRC64;

Query Match 18.8%; Score 521; DB 1; Length 768;
Best Local Similarity 30.5%; Pred. No. 6.4e-27;
Matches 139; Conservative 80; Mismatches 177; Indels 60; Gaps 15;

QY 1 MSRRSLALSGLVCTGLANYSKRAITDSWLSNATYARTRALINLIGADGAWSGADSGI 60
DB 314 VQLRDVVLMNGTVYD-SNGAMDSPLLEWLRQKKVSIIRIFENIGPSAVYPS-ILPGV 371
QY 61 VVASPTDNPYFYTWTRDGLVLTVDLFRNGDTLSLTENYISAQAIIVQGISNPSG 120
DB 372 VIASPSQTHDYFYQWRDASLTINSIVS--HSDPA-ITELQYLVNVRHLQRTNN--- 425
QY 121 DLSSGAG-----LGEPRNVDFTAYTSGWRQRDGPALRATAMIGFQWLLDNG--- 170
DB 426 TLGAGIGYTDVVALGDPKNVNDTAFTEPRGPRQDGPALRSIALIKIIDIYIKOSGTDL 485
QY 171 -----YTSTATDVMPLVRNDLSVAQYMNQYGVDLMEVNGSSFFLIANQHRALVGS 225
DB 486 GAKYPPQSTA-DIFDDIVRMDLRFIIIDHMNSGGLDMEVNGHFFTLVQLSAVDRSL 544
QY 226 FATVAGSSGWCMD--SQAPEILCYL---QSFWTGSFIILANF-----DSSRSGKDA 270
DB 545 YFNASESSPVEBELRQTRDISKEFLVDPNANGFINCKY--NYVETPMADTIRSGLDI 601
QY 271 NTLLGSIHTFDPEAACDSTFQPCSPRALANKEVDSFRSIYTLNDGLSDSEAVANGRY 330
DB 602 STLLAANTVADPSA-SHLPRDIDDPVAVLNTLHMLHMRISYIPINDSSKNATGIALGRY 660
QY 331 PBDTY-----YNGPMPWLCTLAABQYDALYQMDKQSGLEVTVDISDFEALYSDAATG 385
DB 661 PEDYDGYGVGEGNPWLATCAASTTLYOLYRHSQHDLVVPMNDCSNAPWSLWFS 720
QY 386 TYS8-----SSSTYSSIVDAVKTFADGFV 409
DB 721 NLTLLGNDEGYLLIEFNTPAFNQTIQKIFQLADSF 756

RESULT 15
YGI0 METJA STANDARD; PRT; 615 AA.
ID YGI0 METJA
AC Q59005;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical glycosyl hydrolase MJ1610 (EC 3.2.1.-).
GN MJ1610.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_Taxid=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sultion G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Coton M.D., Roberts K.M., Hurre M.A., Kane B.P., Bordovsky M.,
RA Kleen H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii";
RL Science 273:1058-1073(1996).
CC -I- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U67601; AAB99630.1; -.
DR PIR; A64501; A64501.
DR TIGR; MJ1610.
DR Interpro; IPR000165; Glyco_hydro_15.
DR Interpro; IPR008928; Glyco_trans_6np.
DR Interpro; IPR006465; Oligosac_amy1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR TIGRFams; TIGR01577; oligosac_amy1; 1.
DR PROSITE; PS00820; GLUCOAMYLASE; 1.
KW Hypothetical protein; Hydrolase; Glycosidase; Complete proteome.
FT ACT SITE 406 406 GENERAL ACID CATALYST (BY SIMILARITY).
FT ACT SITE 407 407 INTERACT WITH SUBSTRATES (BY SIMILARITY).
SQ SEQUENCE 615 AA; 72007 MW; 2B37EB89F0357BE5 CRC64;

Query Match 8.0%; Score 221.5; DB 1; Length 615;
Best Local Similarity 23.1%; Pred. No. 2.9e-07;
Matches 101; Conservative 46; Mismatches 125; Indels 165; Gaps 20;

QY 59 GIIVASPTDNPYFYTWTRDGLVLTVDLFRNGDTLSLTENYISAQAIIVQGISNP 118
DB 284 GGIAPSL-HPRYRYWGRD-GSYISALDLF-----GIRNPDPFERMSKIQ----- 331
QY 119 SGLSSGAGIGEPKFNVDETAYTGSW-----GRQRDGPALRATAM---IGFQWLL 167
DB 332 -----NAD-----GSLQYVYVNGK-----RLTAIQTDQIGSLIMAM 364
QY 168 DNGYSTATDVMPLVRNDLSVAQYMNQY-----YDLMEVNGSSRF 211
DB 365 DVHYRLT-----GDRKFEYRYWNTTEKANYLRVALNFTPCFDLMEERFGVFAY 414
QY 212 TLIVQHRALVESGAFATV-----GSSCSGCDSPAPEILCYLQSFWTGSFIILANFP 262
DB 415 TWGATYAGLKCAVSMSKAVKRDYKDWGKTIFFLGHVPKR-YLE-----D 461
QY 263 SSSRGDANTLLGSIHTFDPEAACDSTFQPCSPRALANKEVDSFRSIYTLNDGLSDS 322
DB 462 EERFAKSIINPL-----DKTIDTSILGLSYPNLID-----VDBERMIKT-----A 501
QY 323 EAV-----AVGRPEPTYNGNPWLCTLAAB---EQYDALYQMDKQSGLEVTVD 370
DB 502 EALEKAFKYGKVGIGRPBDIYFGNPMWITTLMLLYRRLKVKKEKDQNA---DI 557
QY 371 SLFPFALYSDAATGTYSSSSSYSSIVDAVKTFADGFVSVETHAASNGMSGOYDKSD 430
DB 558 YLQSKKLFWVMWKYSF-----DGLPFEQIHKE 586
QY 431 GEOLASADLTWSYAALL 447
DB 587 GVPMSAMPFGMSNMF 603

RESULT 16

QY 209 SFETIAVHRAALVEGSAFATAVAGSGCWDSQAPILLCYLOSFWTGSFILANFDSSRSRGK 268
 DB 197 SFSSTLTNTSMIPSSSSFTTTTCS-----PYNTSSFLPSSVSSASLS 240
 QY 269 DANTLIGSIHTFPEACDSDSTFQPCSPRALANHKEVDSFRSIYTLNDGLSDSEAVAVG 328
 DB 241 SSSVLEPSTSIIT-----STSTPVTSSSSS-----LSSFTSYSTNLTLTSTSTTTT 284
 QY 329 RYEDDYVYNGNPMFLCTLAAAEQLYDALYQMDQKQSLFETVDVLDLPPKALYSAAAGT-- 386
 DB 285 ---GSATVSSSPFPYSSSVIPISVPSSVSFTSSSSSYTTTLTASNTSVTYTGTSAT 341
 QY 387 -----YSSSSSTYSIYDAVAKTFADGFVSIETHAASNGSSEQ-----425
 DB 342 FTSSPPFYSSSVIPISVPSSVSFTSSNSYTTTLTASNTSIYTGTSATFTSSPP 401
 QY 426 -YDKSD-----GEOLASARDLWYSYALITANNRRNSVVPASWGETTSASVP-----GT 472
 DB 402 FYENSSVIFPISVPSSVSFTSSNSYTTTLTASNTTYTFTGTGTSATFTSSPPYSSNS 461
 QY 473 CATSAIGTYSSVTV--TSMPSIVATGCTTTATPTGSGSVTSTKTATASKT-----ST 526
 DB 462 VIPTSAFSSVSFTSSNSYTTTLTASNTTYTFTGTGTSATFTSSPPYSSNSIIVPTT 521
 QY 527 TTRSG 531
 DB 522 VRTSG 526

RESULT 19

YS89 CAEEL STANDARD; PRT; 3178 AA.

AC 009674; 009625; 0969DA;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypochemical protein ZK945.9 in chromosome II.
 GN ZK945.9/ZK945.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;
 OC Rhabdilitidae; Pelodierinae; Caenorhabdilitis.
 OC NCBI_TaxID=6239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Briscot N2;
 RA Wilkinson-Sproat J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP REVISIONS.
 RA Durbin R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Contains 1 GFS domain.
 CC -1- SIMILARITY: Contains 1 PLAT domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; Z48544; CAB70192.1; -
 DR EMBL; Z48582; CAB70192.1; JOINED.
 DR EMBL; Z48582; CAB70201.1; -
 DR EMBL; Z48544; CAB70201.1; JOINED.
 DR WormRep; ZK945.9; CE25697
 DR InterPro; IPR002111; Cat_channel_TripL.
 DR InterPro; IPR005821; Ion_trans.
 DR InterPro; IPR001024; Lipoxxygenase_LH2.
 DR InterPro; IPR003915; PKD 2.
 DR InterPro; IPR000203; PKD_cys_rich.

DR InterPro; IPR008976; PLAT_LH2.
 DR Pfam; PR01825; GFS; 1.
 DR Pfam; PR00520; ion_trans; 1.
 DR Pfam; PR01477; PLAT; 1.
 DR PRINTS; PR01433; POLYCYSTIN2.
 DR SMART; SM00303; GFS; 1.
 DR SMART; SM00308; LH2; 1.
 DR PROSITE; PS50095; PLAT; 1.
 DR Hypochemical protein; Transmembrane.
 KW TRANSMEM 13
 FT TRANSMEM 13 30 POTENTIAL.
 FT TRANSMEM 51 73 POTENTIAL.
 FT TRANSMEM 2139 2161 POTENTIAL.
 FT TRANSMEM 2348 2367 POTENTIAL.
 FT TRANSMEM 2390 2412 POTENTIAL.
 FT TRANSMEM 2451 2468 POTENTIAL.
 FT TRANSMEM 2483 2505 POTENTIAL.
 FT TRANSMEM 2567 2589 POTENTIAL.
 FT TRANSMEM 2836 2858 POTENTIAL.
 FT TRANSMEM 2939 2961 POTENTIAL.
 FT TRANSMEM 2976 2998 POTENTIAL.
 FT TRANSMEM 3038 3060 POTENTIAL.
 FT TRANSMEM 266 1196 SER/THR-RICH.
 FT DOMAIN 1105 1241 GLY/SER-RICH.
 FT DOMAIN 2071 2120 GFS.
 FT DOMAIN 2182 2305 PLAT.
 SQ SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;

Query Match 5.7%; Score 158; DB 1; Length 3178;
 Best local Similarity 20.1%; Pred. No. 0.041;
 Matches 108; Conservative 81; Mismatches 215; Indels 134; Gaps 19;

QY 32 SNEATVARTALININIGADAWVSGADGIIVASPSPT-----DNPQVFTYTRSGVLVKT 86
 DB 508 STPSSSTASSSVSTASTGTSSTSTQSSSTTYKSETTSDGNTDPDFEAKATITTYDST 567
 QY 87 LVDLFENCDTSLIS-----TIENISAQAIYQGISNPSGDLSSGAGLGPFRNYD 136
 DB 568 SVALTLNLSGLIIGYQTSIECTSPSSNVST-----TKD 602
 QY 137 ETAYTSGWCRPQDGPALPATAMIGQWLLDNGYTSNTDVIPLVRNDLS-----188
 DB 603 GACFTKSVMP-RLGGTYPASTFVPG-----NYTFRAT-----MTDDKYYTYTANY 650
 QY 189 YVAQVWNGQGYDMEBVGNSPFTIAVGHRAALVEGSAFATAVAGSGCWDSQAPILLCYL 248
 DB 651 YIOEYSSIT-----ISESSTSAVASTSTSTPSTPSSTLS--665
 QY 249 QSFWTGSFILANFDSSRSRGKANTLLGSIHTFPEACDSDT-----FQPCSPRA 298
 DB 686 TSIVTEPSSSTRSDSTTSAGSTTLQESTTISEESTTSSSTTIDTGTSSSPSTTA 745
 QY 299 LANHKEVDSFRSIYTLNDGLSDSEAVAVAGRPEDTYNGNPMFLCTLAAAEQLYDALYO 358
 DB 746 DSTSTLSVQFD--FILDGSLSWNET-----RHNEDE-----SINVPLPTVAITP 788
 QY 359 WKQGSLEVTDVSLDPFKALYSDAATGYSSS--STYSIYDAVAKTFADGFVSIETHA 416
 DB 789 TENSQFEECRNVESTFP-LIESTCLNLSNVLNATYSNIP-----IQPLETEL 838
 QY 417 ASNGSSEQYDSDSGQLASARDLWYSYALITANNRRNSVVPASWGETTSASVPTCAAT 476
 DB 839 VGIGTYEPRINMTDLTMQV-----VSHIFILNVAADS-----TSSEVYSTTSGSSSE 869
 QY 477 SAIGYSSVTVTWSPIVATGCTTTATPTGSGSVTSTKTATASKT-STTRSGMS 533
 DB 890 SAISTSGIESTSTLEASTTDAQSDSTSTSDGTTSDTIDSSNSTSTSDSGLS 947
 RESULT 20
 ID ICEN ERWHE STANDARD; PRT; 1258 AA.
 AC P16239;
 DT 01-APR-1990 (Rel. 14, Created)

```

DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN ice nucleation protein.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_Taxid=549;
RP
RN
RP SEQUENCE FROM N.A.
RC STRAIN=M1;
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RL Gene 85:239-242(1989).
CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M25382; AAA24823.1; -.
CC FPI; J00188; J00188.
CC DR HSSP; P06620; 11NA.
CC DR InterPro; IPR000258; Ice_nucleatn.
CC DR Pfam; PF00818; Ice_nucleation; 65.
CC DR PRINTS; PR00327; ICENUCLEATN.
CC DR PROSITE; PS00314; ICE_NUCLEATION; 45.
CC KM Ice nucleation; Repeat; Outer membrane.
CC FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
CC SQ SEQUENCE 1258 AA; 125084 MW; 590E8A130077BBD4 CRC64;
Query March 5.7%; Score 157; DB 1; Length 1258;
Best Local Similarity 22.1%; Pred. No. 0.014;
Matches 132; Conservative 63; Mismatches 239; Indels 162; Gaps 25;
QY 47 GADGAWVSGADSGIVASPTDNDPYTWTRDSGLVLTVDLFRN-----GDTSL 99
DB 299 GADSSLLAGVSGTGTAGESTGTAGVSTGTAKG-----SDLTAGVSGTGDSS-- 350
QY 100 STTENYISQAIVVIGISNPSG-----DLSSG-----AGIGEPKFN 134
DB 351 SLIAGVSGTGTAGDSSLTAGVSGTGTAKGSDLTAGVSGTGTAGDSSLIAGVSGTGT 410
QY 135 VDETAIVGSMGRPGRDGPALATAMIGGQWLNGTSTATDVLWPLVRNDLSVVAQW 194
DB 411 GEESTGTAGVSGTGT-----TAKGSD-----LTAGVSGTGT-----AGDSSLI 450
QY 195 NOTGYDLMEEVNGSSFTTIAVQHRALVEGSAFATAVSSCSGCDQAPELLCY--LQSF 252
DB 451 ---GYSGTGTAGDESSLTAGVSGTGTAKGSDLTAGVSGTGTAGVSSLIAGVSGTGTAG 507
QY 253 TGSFTILANFDSRSKQKANTLL---GSTHTPDPACDSTFPQCSPPALANHKEVDS- 308
DB 508 YGSLTLAGVSGTGTAKHESDLITGYGSTST---AGANSSLIAGVSGTGTASVNSVLTAG 563
QY 309 FRSLTYLTNDGL-----SDSEAVA--VGRYEDPTYYNG--NWFPLCTLAAAO-- 351
DB 564 YGSLTLAGVSGTGTAGVSGTGTAGDSSLTAGVSGTGTASTYHSSLTAGVSGTGTAREOSV 623

```

```

QY 352 -----LYDALYQMDKGSLEVTVDLDPFALYS 380
DB 624 LTTGYSTGTAGADSSLIAGVSGTGTAGYNSILTAGY-----GSTGTAGEGSDLTAGYGS 678
QY 381 DATGTGVSSSSSTVSSIVDAVKTFFADGPFVSVTTHAASNGSMEOYDKSNG----- 431
DB 679 TSTAGADSSLIAGVSG-----TGTAGYNSILTAGVSGTGTAGEGSDLTAGVSGTGTAGA 732
QY 432 -EQLSA---RDLTWSYAALLTA-----NNRNSVVPASMGETSASSVPGTCAATSAIGT 481
DB 733 DSSLIAGVSGTGTASTYHSSLTAGVSGTGTAREDSVLTGTGSGTSTAG-----ADSSLIAG 787
QY 482 YSSVTVTSWPSIVATG-GTTT-----TPYSGSVTSSTKTTATASKSTTTTRSG 531
DB 788 YGSLTGTAGYNSILTAGVSGTGTAKHESDLITGYGSTGTAGDSSLIAGVSGTGTAG 843
RESULT 21
GUX1_ASPAC STANDARD; PRT; 540 AA.
AC 059843;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulobiohydrolase I) (1,4-
DE beta-cellobiohydrolase I) (Beta-glucanancellulobiohydrolase I).
GN CBHI.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_Taxid=5053;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F-50;
RA Takada G., Kawaguchi T., Sumitani J., Arai M.;
RT "Cloning, nucleotide sequence, and transcriptional analysis of
RT Aspergillus aculeatus No.F-50 cellobiohydrolase I (cbhi) gene.";
RL J. Ferment. Bioeng. 85:1-9(1998).
CC -1- FUNCTION: The biological conversion of cellulose to glucose
CC generally requires three types of hydrolytic enzymes: (1)
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
CC Exocellulobiohydrolases that cut the disaccharide cellobiose from
CC the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
CC glucosidases which hydrolyze the cellobiose and other short cello-
CC oligosaccharides to glucose.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl
CC hydrolases).
CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB002821; BAA25183.1; -.
CC DR HSSP; P00725; 1A26.
CC DR InterPro; IPR000254; CBD_fungal.
CC DR InterPro; IPR008985; ConA_like_1ec_g1.
CC DR InterPro; IPR001722; Glyco_hydro_7.
CC DR Pfam; PF00734; CBM_1; 1.
CC DR Pfam; PF00840; Glyco_hydro_7; 1.
CC DR Prodom; PD001821; CBD_fungal; 1.
CC DR Prodom; PD186135; Glyco_hydro_7; 1.
CC DR SMART; SM00236; fCBD; 1.
CC DR PROSITE; PS00562; CBD_FUNGAL; FALSE_NEG.
CC KM Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.

```

FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 540 EXOGLUCANASE 1.
 FT DOMAIN 23 459 CATALYTIC.
 FT DOMAIN 460 505 LINKER.
 FT DOMAIN 506 540 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT SITE 234 234 NUCLEOPHILE (BY SIMILARITY).
 FT ACT SITE 239 239 PROTON DONOR (BY SIMILARITY).
 FT DISULFID 513 529 BY SIMILARITY.
 FT DISULFID 524 539 BY SIMILARITY.
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 540 AA; 57099 MM; 002DFD28DF194D0 CRC64;

Query Match 5.6%; Score 156.5; DB 1; Length 540;
 Best Local Similarity 23.3%; Pred. No. 0.0048;
 Matches 85; Conservative 52; Mismatches 131; Indels 97; Gaps 19;

QY 223 GSAPATAVAGSSGSCWDSQAPDLICVLO---SFWTGSFLANFDSRSRK-----D 269
 DB 186 GAKYGT-----GYCDSQCPRDLKFDGQANIEGMEPSSIDVNAAGNMGSCCPENDIWE 239
 QY 270 ANTLDSIHITFDEACDDSTFQPCSPRAL-ANHKEVYDSEFRTYTLNDGLDSEAVANG 328
 DB 240 ANS-ISSAFAHP---CDSVQQTCTGDTGCGTYSDDTDY-----SGTCDDPGCDFN 288
 QY 329 RY-----PDITYNGNPMFLCTLAABQLYDALYQMDKQSL-EVTVSLIDFPKAL 378
 DB 289 PYFGNTNFPGRKTYDNKRFVYVTFQTHDGTDT-----GTLTEIRLRYQNGVVI 341
 QY 379 YSDAATGYSSSSSYSIYDAVKTADGFSIVETHAASNGMSQYDK-SQGEQI--- 434
 DB 342 GNGPSYTYAASNSITSPCKAEKTLF-GDTNVEETH---GGLSANGDALGQOMVLVS 396
 QY 435 ---SARDLTWSAALLTANNRRNSVVPASWGETSASVP---GTCAT------ 476
 DB 397 LMDHDHADMM-----LSDYPTTSCASSPGVARGTCPTTGNATYVEAN 441
 QY 477 --SAIGYGV---TVTSMPSIVAGGTITAT--PTGSGSVTSTKTTASTKTTT 529
 DB 442 YPMSYTYTSIKRGTLANSTYSGTSSGSSSSTLTITKASTTSSKTTTSTTSS 501
 QY 530 SGMSL 534
 DB 502 SSTNV 506

RESULT 22
 ICEK_PSESX STANDARD; PRT; 1148 AA.

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN INAK.
 OS Pseudomonas syringae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=317;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KCTC 1832;
 RA Jung H.-C., Pan J.-G.;
 Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RL FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
 CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

DR EMBL; AF013159; AAB66891.1; -
 DR HSSP; P06620; 11NA.
 DR InterPro; IPR000258; Ice_nucleatn.
 DR Pfam; PF00818; Ice_nucleation; 57.
 DR PRINTS; PR00327; ICENUCLEATN.
 DR PROSITE; PS00314; ICE_NUCLEATION; 38.
 DR Ice_nucleation; Repeat; Outer membrane.
 FT DOMAIN 161 164 POLY-ALA.
 FT SEQUENCE 1148 AA; 113811 MM; BA4019CF20FBE224 CRC64;

Query Match 5.6%; Score 155.5; DB 1; Length 1148;
 Best Local Similarity 23.6%; Pred. No. 0.015;
 Matches 137; Conservative 66; Mismatches 245; Indels 133; Gaps 27;

QY 32 SNEATVARTAILNNICADG-----AWSGADSGIVASPDNDP-----YRYTWT 77
 DB 200 SNETAGNHSDLLAGYSTGTAGYSTGTSGEDSLTAGYSTQTQAOEGSNLTAGYSTGT 259
 QY 78 --RDSGLVLTVDLFRNGDTSLSTIENYISQAIVQGISNDSGLSSG-----A 126
 DB 260 AGSDSSLLAGYSTGTSGDSSSLTA--GYGSTQTQAOEGSNLTAGYSTGTAGVDSLLA 316
 QY 127 GLGEPEFVNDERTAYTSGWGPQRDGPALBATAMIGFGWLLDNGYSTATDIWPLVRND 186
 DB 317 GYSTGTSGSDSLTAGYSTGT-----TAGEGSN--LTAGYSTGT-----AGSD 359
 QY 187 LSYVAQYMNQGYDLMEVNGSSFTTAV-QHRALVEGSAFATAVSSS-CSWCDGQAPET 244
 DB 360 SSLIA-----GYGSTGTSGSDSLTAGYSTQTQAOEGSNLTAGYSTGTAGVDSLLA 413
 QY 245 LCTLGFPWFGSFLANFDSRSRSGKANTL--LGSHTFPEACDSTQPC-SPRALA 300
 DB 414 YGSTGTSGSDSLTAGYSTGTQAOEGSNLTAGYSTGT-----AGADSLIAGYSTGTSG 469
 QY 301 NHKEVVDSPRSITLNDGL-----SDSEAVAVGRPEPTYNGNPMFLCTLA 347
 DB 470 SSSSLTAGYSTGTAREGSTLTAGYSTGTAGADSSLLA-GYSTGT-----S 516
 QY 348 AAEQLYDALY---QMDKQGLEVTVDVSLDFKALYSDAATGYSSSSSYSIYDA--VK 402
 DB 517 GSSESLTAGYSTQTQAOQSVLTSGYSTQTAGAAANLTGTYSTGTAGHESFLIAGYS 576
 QY 403 TPADFVSVI---ETHAASNGMSQYDKSGEQLSARDL-----TWYALALTA 449
 DB 577 TQTAGHKSILTAGYSTGTARDSYLIAGYSGTGTAGSGSSSLIAGYSTGTATYRSMTLTA 636
 QY 450 -----NNRRNSVVPASWGETSASVPGTCANASALGYSSVVTSMPSIYVAG-GTTT 502
 DB 637 GYSTGTAREHSDLVLTGYGSTTAG-----SNSSLIAGYSTGTAGKSLITAGYSTGT 691
 QY 503 ATP-----TSGSGSVTSTS-----KTTATASTKSTTT 528
 DB 692 AGRSDVLVAGYGS-TSTAGYSSSLIAGYSTGTAGYESTIT 731

RESULT 23
 ICEA_PANAN STANDARD; PRT; 1322 AA.

DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein inaA.
 GN INAA.
 GN -----

```

OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
RT to those of Pseudomonas species and regions required for ice
RT nucleation activity."
RL FEBS Lett. 258:297-300(1989).
CC -|- FUNCTION: ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -|- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -|- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -|- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL, outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17316; CA35194.1; -.
DR PIR: S07053; S07053.
DR HSSP: P06620; IINA.
DR InterPro: IPR000258; Ice_nucleatn.
DR Pfam: PF00818; Ice_nucleation_69.
DR PRINTS: PR00327; ICENUCLEATN.
DR PROSITE: PS00314; ICE_NUCLEATION; 49.
DR Ice_nucleation; Repeat; Outer membrane.
KW DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
FT DOMAIN 162 1281
SQ SEQUENCE 1322 AA; 131094 MW; 89B0E24A837039 CRC64;

Query Match 5.6%; Score 155; DB 1; Length 1322;
Best Local Similarity 22.0%; Pred. No. 0.02; Indels 154; Gaps 27;
Matches 125; Conservative 69; Mismatches 220;

QY 47 GADGAWVSGADSGIVASPTDNDPYFTWTRDGLVLTVDLFRN-----GDTSL 99
DB 315 GADSSLAGVSGTQTAGEESTQTGTAGVSTQTAKG-----SDLTAGVSGTGTAGDS-- 366
QY 100 STTENYISAQAIYVGISNPG-----DISG-----AGIGEPKFN 134
DB 367 SLIAGVSGTGTAGDSSLTAGVSGTQTAKGSDLTAGVSGTGTSGADSSLIAGVSGTQT 426
QY 135 VDETAIVGSGRPPORDPALRATMIGPGQWLNDGTSTATDVTWVLVNDISYVAQYW 194
DB 427 GEESTQTGTAGVSGT-----TAKGSD-----LTAGVSGT-----AGDSSLIAGV 469
QY 195 N-----QTGYDLMEEVSSPFTTIAVORALVEGSAFATAVSGSCMCDSQLPELICYLS 250
DB 470 STQAQKGSGL-----TAGVSGTGTAGYSSLIAG-----YGST-----QT 505
QY 251 FWTGSGFLANPDSRSGKADANTLL--GSHTFPDEACDDSTFQPCSPALANHKVVD 307
DB 506 AGVSGSTLTAGVSGTQTQNESDLITTYGSGTST-----AGANSSLIAGVSGTQTASYNGLT 561
QY 308 S-FRSITLND-----GLSDSEAVAGRYPPEDTYNGPWFLC-----TLAAE 350
DB 562 AGVSGTGTABEGSDLTAGVSGTGTAGSDSSLIAGVSGTGTAGASSLIAGVSGTGTAGYN 621
QY 351 QLYDALYQMDKQSGLEVTVDVSLDFPKALYSDPATGTSSSSSTYSIVDAKTPADGFVS 410
DB 622 SILTAGI-----GSTQTABEGSDLTAGVSGTGTAGASSLIAGVSG-----TQTAGYNS 670
QY 411 IV-----ETHAASNGS--MSEQYDKSD-----GEQLSA--RDLTWSYALLLTA 449

```

```

DB 671 ILTAGVSGTGTABEGSDLTAGVSGTGTAGDSSLIAGVSGTQTASVHSSLIAGVSGTQT 730
QY 450 NNRNNSVVPASMGTSASVPGTCAATSAIGTYSVTWSPSIYATG-GTTTVA----- 503
DB 731 --QEGSVLTITGVSGTSTAG-----ADSSLIAGVSGTGTAGYNSILTAGVSGTGTAGERSD 783
QY 504 TPTGSGSVTSTKTTATATASKTSTTTTNSG 531
DB 784 LTTGSGSTGTAGDSSLIAGVSGTGTAG 811

RESULT 24
ICEN_PSESY STANDARD; PRT; 1200 AA.
AC P06620;
DT 01-JUN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ice_nucleation protein.
GN INAZ.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5203;
RA Green R.L., Warren G.J.;
RT "Physical and functional repetition in a bacterial ice nucleation
RT gene."
RL Nature 317:645-648(1985).
RN [2]
RP SUBCELLULAR LOCATION.
RX MEDLINE=92297969; PubMed=2520825;
RA Lindow S.E., Lahue E., Govindarajan A.G., Panopoulos N.J., Gies D.;
RT "Localization of ice nucleation activity and the IceC gene product in
RT Pseudomonas syringae and Escherichia coli."
RL Mol. Microbe Interact. 2:262-272(1989).
RN [3]
RP 3D-STRUCTURE MODELING OF 490-535.
RX MEDLINE=93360260; PubMed=8355267;
RA Kajava A.V., Lindow S.E.;
RT "A model of the three-dimensional structure of ice nucleation
RT proteins."
RL J. Mol. Biol. 232:709-717(1993).
CC -|- FUNCTION: ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -|- SUBCELLULAR LOCATION: Outer membrane.
CC -|- DOMAIN: CONTAINS 122 IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -|- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -|- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL, outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03035; CA26837.1; -.
DR PIR: A24405; SNPSO.
DR PDB: IINA; 3I-OCT-93.
DR InterPro: IPR000258; Ice_nucleatn.
DR Pfam: PF00818; Ice_nucleation; 61.
DR PRINTS: PR00327; ICENUCLEATN.
DR PROSITE: PS00314; ICE_NUCLEATION; 40.

```

KW Ice nucleation; Repeat; Outer membrane; 3D-structure.
 FT DOMAIN 176 1151 OCTAPEPTIDE PERIODICITY.
 FT STRAND 492 495
 FT TURN 496 497
 FT STRAND 498 501
 FT STRAND 507 511
 FT TURN 512 513
 FT STRAND 514 518
 FT STRAND 524 527
 FT TURN 528 529
 FT STRAND 530 533
 SQ SEQUENCE 1200 AA; 118587 MW; E3BEAD4BEA67ABD0 CRC64;
 Query Match 5.6%; Score 154.5; DB 1; Length 1200;
 Best Local Similarity 22.9%; Pred. No. 0.019;
 Matches 144; Conservative 69; Mismatches 254; Indels 163; Gaps 29;
 QY 9 LSGLVCTGLANVYSKRAITLDSWISNEATVAFATILNNIGADGAMVSGADSGIVVAPSTD 68
 DB 207 IAGYSGTGTAG-----SDSWL-----VAGYSTQTAGDSALTLTGYSGTGTARBSNL 254
 QY 69 NBDYFTTW--RDSGLVLTVDLFRNCDTSLSTIENTISAQAIYQGISNPGDLSG-- 125
 DB 255 TAGYSGTGTAGSDSSLIAGYSTGTSGSDSLTA--GYGSTQTAGSGSNLTAGYSTGT 311
 QY 126 -----AGLSEPKFNVDETAYTSGMGRPO--RDGPALRA-----TAMIGFG 163
 DB 312 AGSDSLIAGYSTGTSGSDSLTLAGYSTGTQAGBSNLTAGYSTGTAGVDSLIAGTG 371
 QY 164 QWL-----LDNGYSTATDVIWPLVNDLSVYAQYMNQGYD 200
 DB 372 STQSGSDALTAGYSTGTQAGBSNLTAGYSTGT-----AGSDSLIA-----GYG 419
 QY 201 LMEVNGSSFTTAV-QHRAIVEGSAFATAVGS--CSKCDSOAPELLCTLGSTWTSFTL 258
 DB 420 STQTSGSDSLTAGYSTGTQAGBSLITLAGYSTGTAGVDSLIAGYSTGTSGSDSLT 479
 QY 259 ANFDSRSRGKANTL---LGSITFPPEACDSTPOPC-SPRALANHKVNDSPRSIYT 314
 DB 480 AGYSGTGTAGBSNLTAGYSTGT-----AGBSLITAGYSTGTSGSBSLITAGYSTGT 535
 QY 315 LNDGL-----SDSEAVNAVGRYPEDTYNGNFWFLCTTAAABQLYDALY---Q 358
 DB 536 AREGSLTAGYSTGTAGADSSLIA-GYGSTGT-----SGSESILTAGYSTGT 582
 QY 359 WDKQGLEVTVDVSLDFKALYSDAATGYSSSSSTYSYDA--VTFADGFVSIYETHA 416
 DB 583 TAQQGSVLTSYGSGSTGTAGAAASNLTTGYSGTGTAGHESFLIAGYSTGTAGHKSIL- 639
 QY 417 ASNGSMSEGYDKSD-----GEQLSA---RDLTWSYALLTLA-----NNRN 454
 DB 640 AGYSGTGTARBSDLIAGYSTGTAGSSSLIAGYSTGTASTRSLTLAGYSTGTARH 699
 QY 455 SVVPASWGBTSASVPGTCAATSAIGTYSVVTWSPSIVATG-GTTTATP-----TGS 508
 DB 700 SGLVGTGYSGTGTAG-----SNSSLIAGYSTGTAGAKRSILTAGYSTGTQAQERTSLVAGY 754
 QY 509 GSVTSTST-----KTTATASKTSTTT 528
 DB 755 GS-TSTAGYSSSLIAGYSTGTAGYESTTLT 783
 RESULT 25
 ICEN_PANAN STANDARD; PRT; 1034 AA.
 AC 047879;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein inau.
 GN INAU.
 OS Pantoea ananae (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=553;
 RN (1) -
 RP SEQUENCE FROM N.A.
 RC STRAIN=KUN-3;
 RX MEDLINE=94264407; PubMed=7764866;
 RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
 RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
 uredovora";
 RL BioSci. Biotechnol. Biochem. 58:762-764(1994).
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
 CC PERIODICITY IS SUPERIMPOSED.
 CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D14992; BAA03636.1; -
 CC PIR; JC2143; JC2143.
 CC DR HSSP; P06620; 11NA.
 CC DR InterPro; IPR000258; Ice_nucleatn.
 CC DR Pfam; PF00818; Ice_nucleatn.
 CC DR PRINTS; PR00327; ICENUCLEATN.
 CC DR PROSITE; PS00314; ICE_NUCLEATION; 34.
 CC KW Ice nucleation; Repeat; Outer membrane.
 CC FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
 CC SQ SEQUENCE 1034 AA; 103378 MW; FA22252D333EADD CRC64;
 Query Match 5.5%; Score 151.5; DB 1; Length 1034;
 Best Local Similarity 21.9%; Pred. No. 0.024;
 Matches 127; Conservative 60; Mismatches 210; Indels 183; Gaps 25;
 QY 47 GADGAMVSGADSGIVVAPSTDNPDYFTWTDGSLVLTVDLFRN-----GTSILL 99
 DB 299 GADSSLIAGYSTGTQAGBSSTGTAGYSTGTQAGK-----SDLTAGYSTGTAGDS-- 350
 QY 100 STIENTYISAQAIYQGISNPG-----DLSSG-----AGLSEPKFN 134
 DB 351 SLIAGYSTGTQAGBSLITLAGYSTGTQAGKSDLTLAGYSTGTAGADSSLIAGYSTGT 410
 QY 135 VDETAYTSGMGRPORDGPALRTAMIGFGQWLLDNGYSTATDVIWPLVNDLSVYAQY 194
 DB 411 GESSTGTAGYSTGT-----TAQKSD---LTAGYSTGT-----AGDSSLIA--- 450
 QY 195 NOTGVDLMEVNGSSFTTAVQHRAIVEGSAFATAVAGSSCSKCDSOAPELLCTY--LQSTW 252
 DB 451 ---GYGSTGTAGBSLITLAGYSTGTQAGKSDLTLAGYSTGTAGYSSSLIAGYSTGT 507
 QY 253 TGSFILANFDSRSRGKANTL---GSITFPPEACDSTPOPCSPRALANHKVNDSP 308
 DB 508 YGSLTLTAGYSTGTQAGNESDLITGYSTGT-----AGANSLLIAGYSTGTASTNSVLTAG 563
 QY 369 DVSIDDFKALYSDAATGYSSSSSTYSS--YDAVKTFPADGFVSIYETHAASNGSSEGY 426
 DB 587 QEN-----SDLTGTGYSTGTAGDSSLIAGYSTGTAGYHSLI---TAGYSTGT 635
 QY 427 DKSD-----GEQLSAR---DLTWSYALL 446

```

Db      636 ERSDDLTTGCGTSTAGADSSLIAGYSGTQTAGYNSILTAGYSGTQTQAQNSDLTTGGST 695
Qy      447 LTNANRRNSVVPASWGETSASVPGTCAA-----TSAIGTSSVTATSW-PSIV 494
Db      696 STAGYESSLI--AGYSGTQTASFKSTILMAGYSSQTAREOSSLTAGYSGTSMAGYDSSL 753
Qy      495 ATGGTTTAA-----TPTGSGSV-TSTSTKTTATASKTSTTT 528
Db      754 AGYSGTQTAGYSGTTLTAGYSGTQTAEHSSLTLAGYSGTAT 793

RESULT 26
MSB2_YEAST
ID      MSB2_YEAST      STANDARD;      PRT; 1306 AA.
AC      P32334;
DT      01-OCT-1993 (Rel. 27, Created)
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      MSB2 protein (Multicopy suppressor of bud emergence 2).
GN      MSB2 OR YGR014W.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      (1)
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92383951; PubMed=1514328;
RA      Bender A., Pringle J.R.;
RT      "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
RL      defect."
RT      Yeast 8:315-323(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288C;
RX      MEDLINE=97435481; PubMed=9290212;
RA      Rieger W., Brueckner M., Schaefer M., Mueller-Auer S.;
RT      "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT      chromosome VII."
RL      Yeast 13:1077-1090(1997).
CC      - PPM: O-glycosylated in the Ser/Thr-rich regions (Probable).
CC      - SIMILARITY: SOME, TO YEAST HKRI.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M77354; AAA34798.1; -
DR      EMBL; Z72799; CAA96997.1; -
DR      PIR; S25370; S25370.
DR      Germonline; 141326; -.
DR      SGD; S0003246; MSB2.
DR      GO; GO:0005887; C:integral to plasma membrane; IMP.
DR      GO; GO:0005034; F:osmosensor activity; IMP.
DR      GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. . .; IGI.
DR      GO; GO:0006970; P:response to osmotic stress; IMP.
KW      Transmembrane; Glycoprotein; Repeat.
FT      DOMAIN 698 816
FT      REPEAT 698 714 1
FT      REPEAT 715 731 2
FT      REPEAT 732 748 3
FT      REPEAT 749 765 4
FT      REPEAT 766 782 5
FT      REPEAT 783 799 6
FT      REPEAT 800 816 7
SQ      SEQUENCE 1306 AA; 133114 MW; 67D5D984D5CA4A6D CRC64;

Query Match 5.5%; Score 151.5; DB 1; Length 1306;
Best Local Similarity 22.9%; Pred. No. 0.034;

```

```

Matches 134; Conservative 83; Mismatches 240; Indels 129; Gaps 24;
Qy      25 ATLDMSLNEATVARTAI-LNNIGADG-----WVSGADSGIVVAPSTDNPDYFTWTRD 79
Db      267 APLQTSSESSFTTASALPLVPSSTVDVGSSASPVVMSAAGQLASSTDNPTMSEFTSLT 326
Qy      80 SGLVLTLYDLFENGCTSLISTENTISAOAIYQGISNPSGDLSSAGLGEFPFNVDEN 139
Db      327 S-----TEVD--GSPVVS--STVSAALSAFLDTSTNSPFIYSPSPSVFVPSGSSB--- 373
Qy      140 YTGSMWRPQRDPALATMIGRCWLNDNGYSTATDLYWPLVRNLDLSVQAQYMQTGY 199
Db      374 -----VASSTANVSSSFSDIPPTSTGSGV-----SVAGSASALAF 412
Qy      200 DLMEVNGSS-----FFTAVQHRALVEGSAFATVAGSSGSCWCD-----SQAPEIL 245
Db      413 QSTTEVYGASASTGWSLSTSLTSLQSTTL-DSSLSLSSASSSDLDYGVSTASLPLLS 471
Qy      246 CYIQSIWTGSPFI--ANPDSRSGKD-ANTLGSIHFFPBAACDDSTFOCFSPRALA 300
Db      472 ASFOASTSSSFVSVPVSFVPQSSSDVASTAPSVS-----SSFSYTSIQAGG 522
Qy      301 NHKEVVDPSRSIYTLNDGLSDSEAVAVGRY----PEDTYNGN-----PWFICTIAABOL 352
Db      523 SSMTNPSSSTIYSSSTGSSSESAASTAGATLSSGSSITWAGLQSQPSTSSLSSESQA 582
Qy      353 -----YDALYQMDKQSLQEVTVSLDFEKLAYSDDA-----TGYSSSS 391
Db      583 TSTSAVLASSSVGTSTPYTTAGAGTASASLSTSTAGTSQVSDPTTALQTSFPASS 642
Qy      392 -----STYSIYDAVTFADGFVSIYETIAANGSMSEQYD-KSDDEQLSARD 438
Db      643 TTEGSETSQGFSTSVLVQMPSSISSEFSPSQTTTQNMASASSSQYTTISSTGILSQVSD 702
Qy      439 LTNVSYAALLTNANRRNSV--VPASWGETSASVPGTCAATSAIGTSS-----YTV 487
Db      703 TSVSYT--TSSSVSQVSDPTPVST-TTSSSVSQVSDPTPVSTTSSSVSQVSDPTPVST 758
Qy      488 TSWPSIVATGCTTTTATPTGSGSVSTSKTTATASKTSTTTRSGMS 533
Db      759 TTSSSSVSQVSDPTPVSTTSSSVSQVSDPT-----SVPTSSRSRSVS 800

RESULT 27
MANB CALSA
ID      MANB CALSA      STANDARD;      PRT; 1331 AA.
AC      P22533;
DT      01-AUG-1991 (Rel. 19, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Beta-mannanase/endoglucanase A precursor (includes: Mannan endo-1,4-
DE      beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-
DE      mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)).
GN      MANA.
OS      Caldociellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC      Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC      Caldicellulosiruptor.
OX      NCBI_TaxID=44001;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93119139; PubMed=1476429;
RA      Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT      "The beta-mannanase from 'Caldociellum saccharolyticum' is part of a
RT      multidomain enzyme."
RL      Appl. Environ. Microbiol. 58:3864-3867(1992).
RN      [2]
RP      SEQUENCE OF 1-346 FROM N.A.
RX      MEDLINE=91247819; PubMed=2039230;
RA      Luetthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT      "Cloning, sequence analysis, and expression in Escherichia coli of a
RT      gene coding for a beta-mannanase from the extremely thermophilic
RT      bacterium 'Caldociellum saccharolyticum'."
RL      Appl. Environ. Microbiol. 57:694-700(1991).

```


CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES. THE SECOND MOST ABUNDANT
 CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
 CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.
 CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
 CC linkages in mannans, galactomannans, glucomannans, and
 CC galactoglucomannans.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.
 CC -1- MISCELLANEOUS: This enzyme is most active at pH 6 and 80 degrees
 CC Celsius.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 CC U (FAMILY 44 OF GLYCOSYL HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcement/>).
 CC or send an email to license@ebi.ac.uk.
 CC -----
 CC EMBL: L01257; AAA71887.1; -.
 CC EMBL: M36063; AAA72861.1; -.
 CC PIR: A48954; A48954.
 CC DR HSP: Q06851; INBC.
 CC DR InterPro: IPR001956; CBD_3.
 CC DR InterPro: IPR008965; Cellul bind.
 CC DR InterPro: IPR001547; Glyco_hydro_5.
 CC Pfam: PF00942; CBM_3; 2.
 CC Pfam: PF00150; cellulase; 1.
 CC ProDom: PD001947; CBD_3; 2.
 CC DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC KM Hydroxylase: Glycosidase; Cellulose degradation; Signal;
 CC MW Multifunctional enzyme.
 CC FT SIGNAL 1 41 POTENTIAL.
 CC FT CHAIN 42 1331 BETA-MANNANASE/ENDOGLUCANASE A.
 CC FT DOMAIN 42 1331 CATALYTIC (MANNANASE ACTIVITY).
 CC FT DOMAIN 326 361 PRO/SER/THR-RICH (PT BOX).
 CC FT DOMAIN 362 518 SUBSTRATE-BINDING (POTENTIAL).
 CC FT DOMAIN 519 564 PRO/SER/THR-RICH (PT BOX).
 CC FT DOMAIN 565 720 SUBSTRATE-BINDING (POTENTIAL).
 CC FT DOMAIN 721 780 PRO/SER/THR-RICH (PT BOX).
 CC FT DOMAIN 781 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).
 CC FT ACT_SITE 162 162 PROTON DONOR (BY SIMILARITY).
 CC FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 CC FT CONFLICT 338 338 T->P (IN REF. 2).
 CC FT CONFLICT 340 346 TPPTPT -> RQHQHQ (IN REF. 2).
 CC SQ SEQUENCE 1331 AA; 146892 MW; 1FBCA51BB8BF0E0 CRC64;

 Qy Query March 5.5%; Score 151.5; DB 1; Length 1331;
 Qy Best Local Similarity 21.1%; Pred. No. 0.034;
 Db Matches 102; Conservative 60; Mismatches 161; Indels 161; Gaps 20;

```

Dh 143 EIKSVLDGNEDFVILINIGNEP---YGNNNYQNWVDTKM1KALRDAGFKHTIIVDAPNW 199
Qy 360 DKQSLSEVTVSLDFFQA-----LYSDATGTYSSSSSTYSSIVDAVKTFADGFSVIV- 412
Dh 200 GQDWSNTRNDNAOSIMWADPLRNLVFSIHMYGVY---NTRASKVEEYKFSFVDKGLPLVI 255
Qy 413 --ETHAASNGSMSBOYOKSDOEOLASBDLFW-----SYALLTANRRRNSVVPASMGET 464
Dh 256 GEFHQHTDDGDPDEALVRAKQYKIGLFSKWSGNSSYGYLDMMNNMPPNPTPGQ- 314
Qy 465 SASVPGTCATSAIIGTYSSVTVTSWPSIVATGTTTATPTPGSGVSTSKTATASKT 524
Dh 315 -----WKTNAIGTSGSTPTPTS-----TYPTPTPTPTPTPTVATPTPTPTVST 360
Qy 525 STTT 528
Dh 361 PATS 364

RESULT 28
FIG2_YEAST STANDARD; PRT; 1609 AA.
ID FIG2_YEAST
AC P25653;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Factor induced gene 2.
GN FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92397594; Pubmed=1523889;
RA Wilson C.; Grisanti P.; Frontali L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
RL cerevisiae chromosome III contains two new open reading frames.";
RL Yeast 8:569-575(1992).
CC CC -1- FUNCTION: Required for efficient mating.
CC CC -1- INDUCTION: By mating pheromones.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC CC or send an email to license@ebi.ac.uk).
CC CC -----
DR EMBL; X59720; CAA42254.1; -.
DR PIR; S25345; S25345.
DR GenBank; U00096.1; -.
DR SGD; S0000685; FIG2.
DR GO; GO:0009277; C:cell wall (sensu Fungi); IDA.
DR GO; GO:0000753; P:cellular morphogenesis during conjugation w. .; IMP.
DR GO; GO:0000755; P:cytogenetic; IMP.
SQ SEQUENCE 1609 AA; 166049 MW; 7D66AD7F85A7B852 CRC64;

Query Match 5.4%; Score 150; DB 1; Length 1609;
Best Local Similarity 20.2%; Pred. No. 0.056;
Matches 124; Conservative 92; Mismatches 211; Indels 186; Gaps 32;

Qy 61 VVASPSTNDNPPYFTWTRDGLVYKTVLDFR-----NGDSTLSLTENYISAQIVGQIS 116
Dh 312 VLTNSND-PSLFTTTEYSSTQLSLNRRSKSTVNFTASIASTPPTGDSATSLIPIS 370
Qy 117 NPSGDLSSGAGLGEPRKN-----VDETAYTGS-----WGRPQDGPALRATAMIGFQ 164
Dh 371 SVGSTASSPGISTANSTQGSNSYVEFSTASGSOYQDWSSSSL-PLSGTT----- 421
Qy 165 WLLDN-----GYTSATD-----IVW-PL-----VRDLSVVA 191

```

```
Db 422 WVVINTNTGSGVSTTSPAVSVTATKTVGVTETVTCWCPLOTGKQALGVSSISV 481
QY 192 QYNQGTGYDLMEEVNGSSFFTIAYOHRLVEGSAFATAVGSS--CSWCDSDQAPILCYLQ 249
Db 482 QASSFSQ-----SSILSSNSSTLAAASNNV-----PESTASSSQYQDMSSSLP-----LSQ 528
QY 250 SFWTGSITLANPDSRSKGNANTLGSIRH-----FDP 282
Db 529 TTV-----VVINTNTGSGVSTTSPAVSVTATKTVGVTETVTCWCPLOTGKQALG 584
QY 283 EAACDDSTFQPCSPRALA-NHKEVVD-SFPRSYTLNDGLSDSEAVAVAGRPEDTYNG-- 338
Db 585 STSATGTSPSSSLTIGISTQLSDATFKGTETINHLH-TESTST---TEPTFEGTGS 640
QY 339 NPWFLECT---LAAEQLDALYQMDKQSLVTVDSLDFEKLALYSDAATGYSSS--- 390
Db 641 DSFYLTCTSEVNLASLSSEYFN-FSSSEGSTATINSTVTF-----GSISKPKSTSVN 692
QY 391 -----SSTSSIVAVATFADGFPVSYETHAASNG--SMSEQDKSDGEO----- 433
Db 693 PTEASQHVSSSVNSLTPTNSSTETIIVINIKTSNKNKYSLLTTQTLKTSKQTLVLTST 752
QY 434 -----LSARDLTW-----SYAL-----LTANNRRNSVVP 458
Db 753 VTTVNGAATETWTWCASSIAVTTISIKTLVLTVEVCHSECTPVITSVTATSTIP 812
QY 459 ASWGETSASSVPGTCAATSAIG-TYSSVTVTSWPSIVATGTTTATPTGSGVSTSTKT 517
Db 813 LL-STSSSTVLSSTVEGAKNPASAEVITNT-QVATSEATSTSTQVATSATATASE 868
QY 518 TATRSKSTSTTRS 530
Db 869 SSTTSQVSTASSET 881

RESULT 29
YG46_YEAST
ID YG46_YEAST STANDARD; PRT; 507 AA.
AC P53301;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 52.6 kDa protein in BUB1-HIP1 intergenic region.
GN YG189C OR G7553.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97279231; PubMed=9133739;
RA Arroyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M.,
  Nombela C.
  "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm
  of Saccharomyces cerevisiae chromosome VII."
  Yeast 13:357-363(1997).
RL -1- SIMILARITY: SOME, TO YEAST UTR2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; Z72974; CA97215.1; -
CC EMBL; X99074; CA67525.1; -
CC PIR; S64507; S64507.
CC HSSP; P23904; 1AJK.
CC Germonline; 141501; -
CC SGD; S000342; CRH1.
```

```
DR GO; GO:0009277; C:cell wall (sensu Fungi); IDA.
DR GO; GO:0000131; C:incipient bud site; IDA.
DR InterPro; IPR008985; Cons like 1ec_g1.
DR InterPro; IPR000757; Glyco hydro_16.
DR Pfam; PF00722; Glyco_hydro_16; 1.
KW Hypothetical protein
FT DOMAIN 63
FT DOMAIN 301
FT DOMAIN 345
FT DOMAIN 387
FT DOMAIN 467
SQ SEQUENCE 507 AA; 52757 MW; 7D7B61F57AE942C CRC64;

Query Match 5.3%; Score 147; DB 1; Length 507;
Best local similarity 20.8%; Pred. No. 0.019;
Matches 125; Conservative 75; Mismatches 143; Indels 258; Gaps 29;

QY 17 LANVSKRATLDWLSNEATVARTALINMGADGAVNSGDSGIVVAPSTDPDYFTW 76
Db 6 LITVLSASSLSLTFPAABESTAT-----ADSTTAASSTASCP----- 42
QY 77 TRDSGLVLTVDLFRNG--DTSLSTLENIYSAQAIYQGISNPGDLSGAGLEPKF 133
Db 43 -----LKT-----TGCTPDTAL-----ATSPSDPSSSKMFTDLGHABIKY 80
QY 134 NVDETAYTSGWRPQDGPALRATAMIGFGQWILNDNGYSTATDIWPLVRNDLSYVQY 193
Db 81 GSDGLSWTLA--KRDNPFLKSNFYIMYK-----LEVILKA 115
QY 194 WNOTGYDLMEEVNGSSFFTIAYOHRLVEGSAFATAVGSSCSWCDSDQAPILCYLQ--- 250
Db 116 ANGTGI-----VSSP-----YLSDDL 132
QY 251 -----FWTG--SFILANPDSRS-----GKDANTLGSIHFPPEAACDDSTF- 291
Db 133 DEIDIEWGGDNTPQPSNFFSKGDTTYYDRGFHGVDTPI--DKFNNTLDMAMDKTWY 190
QY 292 -QPCSPPALANHKEVVDSEFRSITVTLNDGLSDSEAVAVAGRPEDTYT-----NGNP--- 340
Db 191 LDGSEYRVLN-----TSSEG-----YFQSPMYLMGMWAGDPPDNA 227
QY 341 -----WFLCTLAABEQL-DALYQMDKQSLVTVDSLDFEKLALYSDAATG--TYSSSS 391
Db 228 AGTIEW--ACGETNYNDAPF-----TMYIEKIVTVDTSGKKYTGDOSS 269
QY 392 STYSSIVDAVKTADGFVSVIVETHAASNGSMSEQYDK-----SDGEQLSARDL---T 440
Db 270 GSWESIE-----ADG-----GSIVGRYQAGDEFAVLNAGSISSSSTSSST 311
QY 441 WSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIGTYSV-----TTSWPSIVA 495
Db 312 VSSASASTVSSSVSSTVSSASSTVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSVSSSV 371
QY 496 TGGTT-----TTATPTGSGVSTSTKTATASKTS-----TTRSGM 532
Db 372 SSVTTSSISSIFEKQSSSSSKKTVASSSTSESISSSTKTPATVSTTRSTVAPPTQSSV 431
QY 533 S 533
Db 432 S 432

RESULT 30
ID ALAI CANAL
AC O13366;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Agglutinin-like protein ALAI precursor (Agglutinin-like adhesin).
GN ALAI OR ALSS.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
```

CC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98053977; PubMed=9393828;
 RA Gaur N.R., Klotz S.A.;
 RT "Expression, cloning, and characterization of a Candida albicans
 gene, ALA1, that confers adherence properties upon Saccharomyces
 cerevisiae for extracellular matrix proteins.";
 RL Infect. Immun. 65:5289-5294(1997).
 CC -1- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
 CC -1- PTM: N-glycosylated and O-glycosylated (Potential).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AF025429; AAB88863.1; -
 DR PIR: T30531; T30531.
 DR InterPro: IPR008440; Candida_ALS; 1.
 DR Pfam: PF05792; Candida_ALS; 1.
 KW Cell adhesion; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 1419 AGGLUTININ-LIKE PROTEIN ALA1.
 FT DOMAIN 399 404 POLY-THR.
 FT DOMAIN 408 418 POLY-THR.
 FT DOMAIN 437 441 POLY-THR.
 FT DOMAIN 673 676 POLY-SER.
 FT DOMAIN 687 690 POLY-SER.
 FT DOMAIN 700 703 POLY-SER.
 FT DOMAIN 719 724 POLY-SER.
 FT DOMAIN 749 752 POLY-SER.
 FT DOMAIN 787 791 POLY-SER.
 FT DOMAIN 869 872 POLY-SER.
 FT DOMAIN 875 883 POLY-SER.
 FT DOMAIN 901 911 POLY-SER.
 FT DOMAIN 1216 1221 POLY-SER.
 FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 919 919 (POTENTIAL).
 FT CARBOHYD 1301 1301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1326 1326 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1419 AA; 149635 MW; 249F33F688A9D5B6 CRC64;
 Query Match 5.1%; Score 142; DB 1; Length 1419;
 Best Local Similarity 19.9%; Pred. No. 0.16;
 Matches 134; Conservative 89; Mismatches 210; Indels 242; Gaps 34;
 QY 26 TLDSWLSNEATVARTALNINIGADGAVSGADGIVVASTPDNYFT--WTRDSGLV 83
 DB 440 TQDFW--SESFTSTTTITNSL-----KGTDS-VIAREP--NHPVTTFEWS-ESFAT 486
 QY 84 LKTLVD-----LPRNGDTSLSTENTYISAQAVQSGN--PSGDLSSGAGIGEE--- 111
 DB 487 TETITKPEGTDSVIAREPNPTVTTFEWSSEIATTEITNGEG--TDSVIAREHPN 544
 QY 132 ----KFNVDETAVTGS--WGRPORD-----GPAIRATAMIGFQOMLLDNGYSTAT 176
 DB 545 TVTTTFWMSYSATTEITITNKPEGTDSVIAREPNPTVTTFE-----W--SSSYATTEI 597
 QY 177 DIYWP-----LYR--NDLSYVAQYVNGYGLMEEVNNGSSFFTTAVQHRALVEGSAFA 227
 DB 598 ITNGPEGTDSVIAREPNPTVTTFEWSSEIATTEITNGEG--TDSVIAREHPN 544
 QY 228 TAVGSSGSCWDSQAPETLCYLGFWTGSFLLANDSRSGKANDTLGSIHTPPPEARACD 287
 DB 657 TALESSDSNITSSSAGOE-----SSSYVQSSSIVGLSSSDIPSSD 697
 QY 288 DSTFQPCSPALANHKEVVDVFRSIVYTLNDGLSDSEAVAGRPEDTYVNGNPFWFLCTLA 347

DB 698 ----WPSSS-----STGLTSSSESTVSSYSD--SSSIESSTLS 731
 QY 348 AAEO-----LYDALYOMKQGS--LEVTVY-----SLD----- 373
 DB 732 SSDRCSSISDITTSFWDSSSDIESTISITWSSSIDAQSHLYQSVNSISTISQELSSSS 791
 QY 374 ----FFRALYSDAATGYVSSSSSTY-----SSIVDAVK 402
 DB 792 EESSTPATDALVSDASSILSSDTSYPSSTISSDDPFHTIAGSDLSISFISTIVE 851
 QY 403 TPADGVSVIVETHAAS--NSMSGEQYKSGDEQLSADLTWSYVALLTANRRNSVYP--- 458
 DB 852 ISSDS--VSILSPDPASSFSSSSILNSDSSSPSSDQSD-----ILTSSSFLLVPSFS 903
 QY 459 -----ASWGETSASVPGTCA-----ATSAIGTY 482
 DB 904 LSSSSSLSTYPRVNSTYTHASESSSSVASSMSASGANDDTYTLSEDTTSSIGTD 963
 QY 483 SSVT-----VTSWPS-----IVATGCTTTATPTGS--GSVTSKTTA 519
 DB 964 SSTVTFGRDNGDGCIVTGMPSSSIDSEQTSDDVTTSSSFVASSTPTSAREGSTRDNPIDS 1023
 QY 520 TASKSTTTTRSGMGL 534
 DB 1024 SQTSSSTKSSSVSV 1038
 RESULT 31
 ID XYND_CELFI STANDARD; PRT; 644 AA;
 AC P54865;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Endo-1,4-beta-xylanase D precursor (EC 3.2.1.8) (Xylanase D) (XYLD).
 OS Cellulomonas faml.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Micrococcineae; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=221;
 RX MEDLINE=9424155; PubMed=8170399;
 RA Millward-Sadler S.J., Poole D.M., Hentisat B., Hazlewood G.P.,
 RA Clarke J.H., Gilbert H.U.;
 RT "Evidence for a general role for high-affinity non-catalytic
 RT cellulose binding domains in microbial plant cell wall hydrolases";
 RL Mol. Microbiol. 11:375-382(1994).
 CC -1- FUNCTION: Endo-acting xylanase which displays no detectable
 CC activity against polysaccharides other than xylan. Hydrolyzes
 CC glucosidic bonds with retention of anomeric configuration.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
 CC linkages in xylans.
 CC -1- PATHWAY: Xylan degradation.
 CC -1- SIMILARITY: Belongs to cellulase family G (family 11 of glycosyl
 CC hydrolases).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: X76729; CAAS4145.1; -
 DR PIR: I40712; I40712.
 DR PDB: 1E5B; 25-MAY-01.
 DR PDB: 1E5C; 25-MAY-01.
 DR PDB: 1HEH; 10-MAY-01.
 DR PDB: 1HEJ; 10-MAY-01.
 DR PDB: 1XBD; 21-JUL-99.

```

DR PDB; 2XBD, 21-JUL-99.
DR InterPro; IPR001919; Bac_cellose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR008985; Cona_like_1ec.g1.
DR InterPro; IPR001137; Glyco_hydro_11.
DR InterPro; IPR002509; Polyac_deacet.
DR Pfam; PR00553; CBM_2; 2.
DR Pfam; PR00457; Glyco_hydro_11; 1.
DR Pfam; PF01522; Polyac_deacet; 1.
DR PRINTS; PR00911; GLHYDRLASE11.
DR SMART; SM00637; CBD_11; 2.
DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.
DR Xylan degradation; Hydrolase; Glycosidase; Signal; Repeat;
KW 3D-structure.
FT SIGNAL 1 43 POTENTIAL.
FT CHAIN 44 644 ENDO-1,4-BETA-XYLANASE D.
FT DOMAIN 44 230 CATALYTIC.
FT DOMAIN 231 245 LINKER ("HINGE") (GLY-RICH BOX).
FT DOMAIN 246 644 2 X 88 AA APPROXIMATE CBD-LIKE REPEATS.
FT REPEAT 246 333 1.
FT REPEAT 557 644 2.
FT DOMAIN 337 350 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 337 350 LINKER ("HINGE") (GLY-RICH BOX).
FT DOMAIN 231 238 POLY-GLY.
FT DOMAIN 241 245 POLY-GLY.
FT DOMAIN 548 558 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 126 126 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 216 216
SQ SEQUENCE 644 AA; 66581 MW; 56B045CCE0E1820 CRC64;

Query Match 5.1%; Score 141.5; DB 1; Length 644;
Best Local Similarity 20.7%; Pred. No. 0.059;
Matches 142; Conservative 76; Mismatches 238; Indels 229; Gaps 35;

QY 28 DSWLSNATYARTRAI--LNNIGADGAWVSGADSGIVVASPS---TDN-----PDYRYT- 75
DB 3 DSEATRTTTRRRRLQALTLGILAAAGALVAGH--LAAASPAAAVTNTTGTHDGYTSF 59
QY 76 WTRDSGLVLTVDLFRNGDTSLSTIENYISAOAIVOGISNPBGDLSGAGLEPRFNV 135
DB 60 WTPSPGVS---SMDLNGGGGYTRMSNTGNFVAG---KGMSTGRTKTVSISG---QFNP 108
QY 136 DERYTYSKWRPQDQALRATAMIGQWMLDNGYSTATD-----TWPLV 183
DB 109 SRNAVLTLYGMTQ--SPLEVEYIVDSWGYRPTGTFMGYTVSDGTVDYRTQKVNPSI 166
QY 184 RNDLSYVAOYWN-----QTG-----YDLWE-----EYNGSSF 211
DB 167 EGDSSSTYQYVAVQQRKTGTTTSGNHFDMASKMNLGRHNYMATRGYQSSGSSI 226
QY 212 TIAVQHRALVEGSAFA---TAVGSSCSWCDSPAPEILCYLOSF-----WTGSF 256
DB 227 TVS-----EGSGGGGGGSDTGGGSGTSCSVTRAPEMSDRNVTVYSVSGSSAWTVNL 279
QY 257 ILANFDSRSGKDKAN-----TLGSIHTF-----DPEACDOSTFPPCS 295
DB 280 ALNGSQTIOASWNAVTVGSGSTRITVTPNGSGNTGVTVMKNGSSTTPAATCAGSGGTAT 339
QY 296 PRALANKEVVDPRSIV--TLNDG-----LSDSEAVAVAGRYBEDITYVNGNP 340
DB 340 PTPPTPTPTPPQSGAGVGLTFPDGPNVTGTTNOILSTLYQGYATP-VFPTGNAQGNP 398
QY 341 WFLCTTAAEQLYD-----ALYQWD-----KQGSLEVTVSILDF 375
DB 399 SLW-----QAYKNAGVQIGNHSDHPLVMNSQSDMQSLTTCQAICQTAGVTPLF 451
QY 376 KALY--SDAATGYSSS-----SSTYSIVDAVKTADGFVSIV- 413
DB 452 RPPYGEENATLRQVESLGLREIIMWDVSDQPMNNAASQIRQASRLTNGQIIMHMPA 511
QY 414 -THAASGMSSEQYDKSDGEOLSARDLTWSYAALLTANNRNSVVPAS-WGETASSVPG 471

```

```

DB 512 ATVOALPGLI-----QDLRSRLCTGHISSTG-----RAVAPSSAGGGGGGGCTG 558
QY 472 TCAATSAIG-----TY-----SSTVYT-----SPSIVATGTTTATPT 506
DB 559 SCGSVAVRGEEMADRFNVTVSVSGSSWVTLGLANGGQSVQSSWMNALTSSTGTVARPN 618
QY 507 GSGS---VT-STSKTATASKSTT 527
DB 619 GSGNSFGVTFPKYKSSATPGATCAT 643

RESULT 32
HKRI_YEAST STANDARD; PRT; 1802 AA.
ID HKRI_YEAST
AC P41809;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hansenula WRAKII killer toxin-resistant protein 1 precursor.
GN HKRI OR YDR420W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YNN 295;
RX MEDLINE=94156857; PubMed=8113191;
RA Kasahara S., Yamada H., Mio T., Shiratori Y., Miyamoto C.,
RA Yabe T., Nakajima T., Ichishima E., Furuchi Y.;
RT "Cloning of the Saccharomyces cerevisiae gene whose overexpression
RT overcomes the effects of Hm-1 killer toxin, which inhibits
RT beta-glucan synthesis."
RT J. Bacteriol. 176:1488-1499 (1994).
CC - FUNCTION: Could regulate beta-glucan synthesis. Overexpression
CC provides resistance to Hm-1 killer toxin.
CC - SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC - PFM: Could be O-glycosylated in the serine/threonine-rich domain.
CC - SIMILARITY: SOME, TO YEAST MSB2.
CC - This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S69101; AAB30051.1; -.
DR GenBank; U0912; -.
DR SGD; S0002828; HKRI.
KW Glycoprotein; Transmembrane; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 1802
FT TRANSMEM 1486 1506
FT DOMAIN 23 1478
FT 453
FT 453
FT REPEAT 453 480
FT REPEAT 481 508
FT REPEAT 509 536
FT REPEAT 537 564
FT REPEAT 565 592
FT REPEAT 593 620
FT REPEAT 621 648
FT REPEAT 649 676
FT REPEAT 677 704
FT REPEAT 705 732
FT REPEAT 733 760
FT REPEAT 761 788
FT CARBOHYD 24 24
N-LINKED (GLCNAC. .) (POTENTIAL).

```

```

FT CARBOHYD 1252 1252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1293 1293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1342 1342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1400 1400 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1802 AA; 188890 MW; E344CA6469785A24 CRC64;

Query Match
Best Local Similarity 21.4%; Score 141; DB 1; Length 1802;
Matches 127; Conservative 78; Mismatches 243; Indels 146; Gaps 25;

QY 15 TGLANISKRATLDSWLSNEATVATFALLNINICADGAWGSGDGI--VVASPST-NDP 71
DB 234 TGFSTNTVS-----FENVEBEYAMSKQLSEYSSSTVSGGESTADKTSSPITSFSSS 289
QY 72 YFTYTW-----TRDSGLVUKTLVDFRNGDITSLSTIE-----NYISQAIV- 112
DB 290 YSCOTTSTFSESRRVAVGVRPSITQTSTIDSFMSSEVELSTYDLSAGNYPDQELIYD 349
QY 113 ----QGISNPGDLSGAGIGEPKFNVDETAYVSGWRPQRDPALRATAMIGQWLLD 168
DB 350 RPATSTASTSSSAGSGVRESNTFAVSSIST----- 382
QY 169 NGYTSTATIV-----WPLVNDLSYVAQVNNQGYDLM--EVNGSSPFTTA 214
DB 383 NFIVSASDPTVSTSTNTVPYSSVHSTFVHATSSSTYISSLYSSPILSASVSHFGVA 442
QY 215 VQHRALVSGAFATVAGSSGCMDS-----QAPELICVLQSFMTGS-----FILA 259
DB 443 PEPASATISFSSVPAVASTYTSSPSASVVPVAVASPPVAVASTYTSSPSAPALIS 502
QY 260 NFDSRSGKADANTLIGSIHTPDEA--ACDDSTP--OPCSPRALA-----NHKEVVDSEFRI 312
DB 503 YTTSSPS---APVAVSSTYTSSPSAPALISSTYTSSPSAPVAVSSTYTSSPSAPALIS 559
QY 313 YTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLYDALYQDKGSLVETDVL 372
DB 560 YL-----SSPSAPAV-----SSTYTSSP-----SAPALISSTYTSSPSAPAVASTYT 603
QY 373 DFFKALYSDAATGTYSS-----SSTYSIYDAKTFADGVSIVETHAANGSMSE 425
DB 604 SSPSA--PALISSTYTSSPSAPVAVSSTYTSSPSAPALISSTYTSSPSAPVAV---VSST 657
QY 426 YKSIDSEQLSARDLTVSYALLTANNRRNSVVPASGCTSSASVPTCAATSAIGTYS-- 483
DB 658 YTSSP-----SAPALISSTYTSSPSVPAVASTYTSS--PAPALISSTYTSSPS 705
QY 484 ----SVTVTSWPSIVATGTTTATPTGSGVTS--TSKTTATASKTSTTTRS 530
DB 706 APVAVSSTYTSSPSAPALISSTYTSSPSAPVAVSSTYTSSPSAPALISSTYTSS 759

RESULT 33
ICEN_XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN Xanthomonas campestris (pv. translucens).
OS Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxId=343;
RN SEQUENCE FROM N.A.
RC STRAIN=X56S;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RL Xanthomonas campestris pv. translucens.";
MO1. Gen. Genet. 223:163-166(1990).
CC -I- FUNCTION: Ice nucleation proteins enable bacteria to nucleate

```

```

CC crystallization in supercooled water.
CC -I- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -I- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -I- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -I- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X52970; CAA37140.1; -.
DR HSSP; P06620; IINA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 81.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 57.
KW Ice_nucleation; Repeat; Outer membrane.
SQ SEQUENCE 1567 AA; 152548 MW; CBB451D959ECAD63 CRC64;

Query Match
Best Local Similarity 23.1%; Score 140.5; DB 1; Length 1567;
Matches 126; Conservative 68; Mismatches 229; Indels 123; Gaps 25;

QY 47 GADGAMV-----SGASGIVVASPSND-----NDPYRYMT--RDGSLVUKTLV 88
DB 725 GADSTLIAGYSTQTSGSDSLTAGYSGTQARKGSDVTAGYSTGTAGADSTLIAGYGS 784
QY 89 DIFRNDGTSLSTIENYISAQIVQGISNPGDLSG-----AGIGEPKFNVDETA 139
DB 785 TQTSQSDSLTA---GYSGTQARKGSDITAGYSGTGTAGADSTLIAGYSGTQTSQSDS 841
QY 140 YTGSGWRPQ--RDGPALRATAMIGQW-----LLDNGYTSTATDIYVPLVNDLSY 190
DB 842 LTAGYSTGTQARKGSDVTA---GYSGTGTAGADSTLIAGYSGTQTR-----AGSDSLR 891
QY 191 AQYWN--QTGYDLMEEVNGSSFTIANQVHALVBSAFATVAGSSCSCMQSQAPELLICY 249
DB 892 AGYSGTQARKGSDVTAGYSGTGTAGADSTLIAGYSGTQTSGS---DSLTAGYSGTQ 946
QY 250 SEWTSGFILANFDSRSGKADANTLIGSIHTPDEACDSTFQPCSPRALANHKEVVDSE 309
DB 947 TARKGSDMTAGYSGTGTAGADSTLIAGYSGTQTSQSDSLTAGYSGTQARKGSDVTAG 1006
QY 310 RSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLYDALY---QNDKQGSLE 366
DB 1007 GS-----TGTAGADSTLIAGYGS-----TQTAGSDSLTAGYSGTQARKGS-- 1048
QY 367 VTDVSLDFPKALYSDAATGTYSSS-----STYSIYDAVTFADGVSIVETHAANG 419
DB 1049 --DVTAGY-----GSTGTAGADSTLIAGYSGTQTSQSDSLTAGYSG---STYTAHQ 1094
QY 420 GS-MSEQYDKSD-----GEOLSA---RDLTWSYVALLTANNRRNSVVPASWGE 463
DB 1095 GSDITAGYSGTGTAGADSLIAGYSGTQARKGSDNLTAGYSGTQA--REDSSLTAGYGS 1152
QY 464 TSASSVPTGCAATSAIGTYSVTVTSWPSIVATG--GTTTATPTGSGVTSISKTTATNS 522
DB 1153 TSTAG-----HDSLIAGYSGTQAGYNSILTTGYSGTQTAQE--SSSLTAGYSGT 1205
QY 523 KTSTTT 528
DB 1206 YDSTLT 1211

RESULT 34

```

```

SP96.DICDI
ID _SP96.DICDI STANDARD; PRT; 600 AA.
AC PI4328;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Spore coat protein Sp96.
GN CoRA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxId=44689; [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=90067962; PubMed=2587278;
RA Fosenau K., Loomis W.F.;
RT "Sequence of the Dictyostelium discoideum spore coat gene SP96." ;
RL Nucleic Acids Res. 17:9489-9489(1989).
CC -I- SUBCELLULAR LOCATION: Outer layer of the coat matrix and
interapore matrix.
CC -I- PTM: PHOSPHORYLATED AND FCOSYLATED. MAY BE PHOSPHOGLYCOSYLATED,
MAY CONTAIN GLCNAC-ALPHA-1-P-SER RESIDUES.
CC -I- SIMILARITY: Contains 4 prespore motif repeats.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb.ch/announce/
or send an email to license@ib.slb.ch).
-----
CC DR EMBL; X16491; CAA34508.1; -.
DR PIR; S07638; S07638.
DR GLCOSuiteDB; PI4328; -.
DR DR DictyBase; DB0185032; CoRA.
DR InterPro; IPR007643; Dict_spore_N.
DR InterPro; IPR003645; FOIN.
DR Pfam; PF04562; Dict_spore_N_1.
DR SMART; SM00274; FOLN; 6.
KW Glycoprotein; Phosphorylation; Repeat; Sporulation.
FT REPEAT 185 197 PRESPORE MOTIF 1.
FT REPEAT 221 233 PRESPORE MOTIF 2.
FT FT REPEAT 298 310 PRESPORE MOTIF 3.
FT FT REPEAT 395 407 PRESPORE MOTIF 4.
SQ SEQUENCE 600 AA; 59589 MW; 616AE6D02B5F1071 CRC64;

Query Match 5.0%; Score 138.5; DB 1; Length 600;
Best Local Similarity 29.3%; Pred. No. 0.085; Indels 1; Gaps 1;
Matches 44; Conservative 32; Mismatches 73;

QY 380 SDATGTGSSSSSTYSIVDVKTPADGFVSIETHAASNGSMSEPDQKDSDOEQLSARDL 439
Db |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
447 SSSAGSGSAASSPSSSAASSPSSSAASSPSSSAASSPSSSAASSPSSSAASSPSSSAASSSA 506
QY 440 TWSYAALLTA-NNRNSNVVPASWCGETSASVPGTCATSAIGTVSVTVTSMPISVATGG 498
Db ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
507 PSSSSSSSSABSSSSSSSSSSSSSSSSSSSSSSSSSATATAATTATTAATTATTTATTTATTA 566
QY 499 TTTTATPTGSGSVSTSKTKTATASKITSTTT 528
Db |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
567 TTTAATTATTTAATTATTTATTTATTTATTTATTT ATT 596

RESULT 35
VG_LX_HSVB STANDARD; PRT; 797 AA.
AC P28968;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN

```

```

OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1)
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1310606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1."
RL Virology 189:304-316 (1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ch/announcement/
CC or send an email to license@isb-ebi.ch).
CC -----
DR EMBL, M86664; AAB02506.1; -.
DR PIR, H36802; VGBEX1.
DR InterPro: IPR007110; Ig-like.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 797 POTENTIAL.
FT DOMAIN 23 465 SER/THR-RICH.
FT TRANSMEM 766 790 POTENTIAL.
FT CARBOHYD 590 590 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 797 AA; 80342 MW; 50C9ED9211FE5B82 CRC64;
Query Match 5.0%; Score 138; DB 1; Length 797;
Best Local Similarity 32.0%; Pred. No. 0.13;
Matches 47; Conservative 21; Mismatches 59; Indels 20; Gaps 4;
OY 400 AVKTEPAQPFVSIVETTHAASNGSMSEQYDKSDGEQLSARDLTWSYALLTANRNSVVA 459
Db 14 AVSLVIALGSTTTTETTTSSSGTSQSGSTGTTNSSSPSTSPPT--TSSPPTSTHTS 71
OY 460 SWGRTS-----ASSVPQTCATSAIGTYSV-TVTSMPISVATGTTTATPTGSG 509
Db 72 SPSSSTGQSSSTATSSASPASTASTSTISPTSTETTTTTPASTTPTTTTAAPTTAA 131
OY 510 SV-----TSTSKTTATASKTSTTT 528
Db 132 TTTAVTTAATSAETTTATATATATSTPT 158
RESULT 36
WSC3_YEAST STANDARD; PRT; 556 AA.
AC Q1215;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cell wall integrity and stress response component 3 precursor.
GN WSC3 OR YOL105C OR HRS56.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96076631; PubMed=7502582;
RA Vandenbol M., Durand P., Portetelle D., Hilger F.;
RT "Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV
RT including the Y1-H3 retrotransposon, the SUI1(+) frameshift
RT suppressor gene for tRNA-Gly, the yeast transfer RNA-Thr-1a and a
RT delta element."
RL Yeast 11:1069-1075 (1995).
CC -----
CC -I- SIMILARITY: Contains 1 WSC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; 248149; CAA8151.1; -
DR EMBL; 274847; CAA9123.1; -
DR PIR; S51892; S51892.
DR GenOnline; 143527; -
DR SCD; S0005465; MSCJ.
DR GO; GO:0004888; F:transmembrane receptor activity; IGI.
DR GO; GO:0009408; P:response to heat; IGI.
DR GO; GO:0007266; P:rho protein signal transduction; IGI.
DR InterPro; IPR002889; MSC.
DR Pfam; PF01822; MSC; 1.
DR SMART; SM00321; MSC; 1.
KW Cell wall; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 38
FT CHAIN 1 39 556
FT DOMAIN 39 132
FT TRANSSEM 137 348
FT TRANSSEM 385 405
FT CARBOHYD 84 84
FT CARBOHYD 367 367
FT CARBOHYD 370 370
FT CARBOHYD 473 473
FT CARBOHYD 480 480
SQ SEQUENCE 556 AA; 58229 MW; DD37E27180001DA CRC64;
Query Match 4.9%; Score 137; DB 1; Length 556;
Best Local Similarity 27.7%; Pred. No. 0.096;
Matches 48; Conservative 41; Mismatches 68; Indels 16; Gaps 5;

QY 371 SLDFPKAL--YSDAATGYSS-----SSSTYSIYDAVKTADGFSIVETHAASNGS 421
DB 93 SVSFLSLTFTSDNSGCTKCSGWPYMGCGSSYTN-----YVNAETPVSSESSSMEGS 148
QY 422 MSQOYKSDGQESADLTMSYALLTANNRNSVVPASNGEFSASSVPTCAATATICT 461
DB 149 -STSYMPSTTSSLSAQI--SSTRTSTDKSEMAIATVSTTSSSTSSSTSS 205
QY 482 YSSVTVMSPIATGTTTATPTGSGVTSKTTATATASKTSTTRSGMSL 534
DB 206 STTSSSTTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 258

RESULT 37
CH12_RHIOL STANDARD; PRT; 542 AA.
AC P29027;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Chitinase 2 precursor (EC 3.2.1.14).
GN CH12
OS Rhizopus oligosporus.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Rhizopus.
OX NCBI_TaxID=4847;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-38.
RC STRAIN=Saio IFO 8631;
RX MEDLINE=93054356; PubMed=1429462;
RA Yanai K., Takaya N., Kojima N., Horiuchi H., Ohta A., Takagi M.;
RT "Purification of two chitinases from Rhizopus oligosporus and
RT isolation and sequencing of the encoding genes";
RU J. Bacteriol. 174:7398-7406(1992).
CC - FUNCTION: Probably involved in the apical growth and branching
CC of fungal hyphae.
CC - CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.

CC - SUBUNIT: Monomer.
CC -1 SUBCELLULAR LOCATION: Secreted (Probable).
CC -1 PTM: O-glycosylated.
CC -1 SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl
CC hydrolases).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D10158; BAA01022.1; -
DR PIR; B47022; B47022.
DR HSSP; P23472; 2HW.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18A.
DR InterPro; IPR005089; Glyco_hydro_18C.
DR Pfam; PF03427; CBM_19; 1.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KW Hydrolase; Glycosidase; Chitin degradation; Signal; Zymogen;
KW Glycoprotein; Chitin-binding.
FT SIGNAL 1 22
FT CHAIN 23 446
FT PROPER 447 542
FT DOMAIN 23 312
FT DOMAIN 313 354
FT DOMAIN 355 406
FT DOMAIN 407 446
FT ACT SITE 166 166
SQ SEQUENCE 542 AA; 56528 MW; 3B8F17DA551FDDA7 CRC64;
Query Match 4.9%; Score 136; DB 1; Length 542;
Best Local Similarity 23.1%; Pred. No. 0.11;
Matches 83; Conservative 50; Mismatches 146; Indels 80; Gaps 17;

QY 224 SAFATVSSGSCWQDAPEILCYL-QSFWTGSFILANFDS-SRSGKANDTLGSIHTFD 261
DB 12 SAEFLASTGVQAW-SHGPNVWYWGQNSGASNTQASIGTCEGQVDAVLSPFLHVN 70
QY 282 ----PE-----AACDDSTPQPCS-----PRALAHKEVDSFRSITLNDIGDSEAVAVGR 329
DB 71 VGGIFPINISSAGAGYFPNTQLSCPAVGADIKKQDGVKVT-----LSLGAAGVYG 125
QY 330 YPEDFYNGNPWFLLCTLAABQLYDALYQWDKQSLVTVDSLDFPKALYSDAATGYSS 389
DB 126 FTSD-----AGQGFAPQRTNMLFFGGNSD-----TRPFDVAVIDGVDLIEGG 168
QY 390 SSSTYSIYDAVK-----TPADGFSIVETHA-----SNGSMSEQYKSDG 431
DB 169 SSTGYAFAFNALROKFSNFIIGAPQCFPPDALLGSVINSASFDYVNVQFYNNYSATG 228
QY 432 EQLSARDLTMSYALLTANNRNSV--VPASNGEFSASSVPTCAAT-----SAIGTY 462
DB 229 SSENFD--TWDMAKTSSPKVKKIMFTVPGSSTAAGSVTPSTLQTTVPISLASYSS 286
QY 483 SSVTV--TSMV-----SIVATGTTTATPTGSGVTSKTTATATASKTSTTT 528
DB 287 GGVSVWDASQANNGSFGNSQLSVHSGG-STPPPPSSSATTKTTTATATASKTSTTT 344

RESULT 38
Y109_YEAST STANDARD; PRT; 995 AA.
ID Y109_YEAST
AC P40442;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 99.7 kDa protein in SDb1 5' region precursor.
GN Y1169C OR Y19402.07C.

```

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=82886 / AB972;
RX MEDLINE=97313266; PubMed=9169870;
RA Churche C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA Harris D.E., Hornsby T., Hunt S., Jags R., Jones M., Lye G.,
RA Moulton S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA Skelton J., Smith V., Walsh S., Whitehead S., Barrett B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
RL Nature 387:84-87(1997).
CC -1- SIMILARITY: Contains 1 methyl-accepting transducer domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 246921; CA87023.1; -.
DR PIR; S50358; S50358.
DR GerMOnline; 139704; -.
DR SGD; S0001431; YII169C.
DR InterPro; IPR004089; Chmtaxie_transd.
DR PROSITE; PS50111; CHEMOTAXIS_TRANSDUC_2; 1.
KW Hypothetical protein; Signal.
FT CHAIN 1 23 POTENTIAL.
FT DOMAIN 24 995 HYPOTHETICAL PROTEIN YII169C.
FT CARBOHYD 26 253 METHYL-ACCEPTING TRANSDUCER.
FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 664 664 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 995 AA; 99735 MW; F63E287A03F137EC CRC64;

Query Match 4.9%; Score 136; DB 1; Length 995;
Best Local Similarity 22.0%; Pred. No. 0.25;
Matches 77; Conservative 61; Mismatches 140; Indels 72; Gaps 8;

QY 188 SYVAQYWNQGYLMEEVNNGSSFFTLAVQHRALVEGSAFATAVGSSGCMCDQAPETLCY 247
DB 20 SALGQYYSN-----ISISNNSST-----SVSSSSGVSISSTIAETSSSATDIL-- 66
QY 248 LQSFMTGSLFANFDSSRSKDKANTLLGSIHTFDPEAACD-DSTFQPCSPRALNHKEV 306
DB 67 --SSITQASASTSGVSSVGPSSSVSSSVSSQSSSVSSQSSASADVSASSVS 124
QY 307 DSRFSIYTLNDGLSDSRAVAVGRYPEDTYNGNPFICLTAAAEQLYALYQMDKQSL 366
DB 125 QASASTSDSVSSVS-----SSSSS 143
QY 367 VTIVSLDFFKALYSDAATGTYSSSSSTYSISYDAVKTFADGFYSIVETHAASNGSMEQY 426
DB 144 ASDVS-----SSVQSSSSASADVSQSSASASADVSQSSASADVSQSSASADVS 133
QY 427 DKSDGEQLSARDLTWVSAAALLTANRRNSVVPASWGETS--ASSVPGTCATSAIGTYS 484
DB 194 SQSSS---SADVSSSVSSQSSSSASDVSSSVSSQSSASTSDVSSSVSSQSSASTSDVSS 250
QY 485 VYTWSMYSIATGTTTATPTGSGVTSKTTATATSKSTTTTSSMSL 534
DB 251 QSVSSASGSSSSPQSTSTASSTASSTASSTASSTASSTASSTASSTASSTASSTAS 300

RESULT 39
GUX1_PENJA STANDARD; PRT; 537 AA.

```

```

AC Q06886;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I)
DE (1,4-beta-cellulobiohydrolase).
GN CBH1.
OS Penicillium janthinellum (Penicillium vitale).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxId=5079;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C41;
RX MEDLINE=93178976; PubMed=8440481;
RA Koch A., Weigel C.T.O., Schulz G.;
RT "Cloning, sequencing, and heterologous expression of a cellulase-
RT encoding cDNA (cbh1) from Penicillium janthinellum.";
RL Gene 124:57-65(1993).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellobiose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl
CC hydrolases).
CC
CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59054; CA41780.1; -.
DR PIR; J00150; J00150.
DR HSSP; P00725; 8CEL.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR008985; ConA like lec_g1.
DR InterPro; IPR001722; Glyco_hydro_7.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; Glyco_hydro_7; 1.
DR SMART; SM00236; fCBP; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; Signal.
FT CHAIN 1 18 POTENTIAL.
FT DOMAIN 19 537 EXOGLUCANASE I.
FT DOMAIN 19 453 CATALYTIC.
FT DOMAIN 454 477 LINKER.
FT ACT_SITE 478 537 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 235 235 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 240 240 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 509 526 BY SIMILARITY.
FT DISULFID 520 536 BY SIMILARITY.
SQ SEQUENCE 537 AA; 56842 MW; A6B9C6B7317BE4 CRC64;

Query Match 4.9%; Score 135.5; DB 1; Length 537;
Best Local Similarity 21.7%; Pred. No. 0.12;
Matches 94; Conservative 57; Mismatches 135; Indels 147; Gaps 23;

QY 182 LVENDLSYVAQYWNQGYLMEV-----NGSSFFTLAVQHRALVE-----GSAR 226
DB 133 LLENDPTTY--QKFNLDNQEFTPDVDSNLPCCGLNGALYFVMDADGCMKPYPTKAGAXY 190
QY 227 ATAVGSSGCMCDQAPETLCYQ-----SFWTGSFLFANFDSSRSK-----DANTL 273
DB 191 GT-----GYCDSQCFRDLKFINGQANVDGWTFSKQDVNSGIGNHSCCAEMDIWANSI 244

```



```

QY 274 LGS1-----HTPPEAACDSTFOPCSPRALANH--- 302
DB 245 SNAVTHPCDTPSCQMTGRCGCTYSTRYGTCPD--GCDNPFR---MGTVMFYGG 299
QY 303 KEVYDPSRSIYTL-----NDGLSDSEAVAVGR-YEDTYTNGNPMFLCTLAAEQLYDA 355
DB 300 GETIDT-KSPFTVVTGFLTJDTGTSTGLSEIKRFYVQGGKVINP----- 343
QY 356 LYQWMDQGLS-----EVDVSLDFPKALYSDAATGYSSSSSYSDAVKTFADGV 409
DB 344 -----QSTIVGVSGNSTDSWCAQKSAFGD-----TNEFSKHGGMAGMAGLADGV 391
QY 410 ----SIVETHAASNGSMSEQYKSDGEQLSARDLTWGYAALLTANNRNSVVPASWGETSA 466
DB 392 LVMSLWDDHAS-----DVLMDSTYPT--NATSTTPGAKRGTCOD 429
QY 467 SSVPTGCAAT--SAICTYSSVTVTSPSIVATGTT---TTATPTGSGSVTSTKTAT 520
DB 430 SRPFTVESTYPAVAVYSNIK-TGPLNSTFTGTGTTSSSTTTTSTKSTSTSSSKTTT 488
QY 521 ASKTSTTTRSGMS 533
DB 489 V--TTTTTSSGSS 499

RESULT 40
PHP_DROME STANDARD; PRT; 1589 AA.
ID PHP_DROME STANDARD; PRT; 1589 AA.
AC P39769;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Polyhomeotic-proximal chromatin protein.
PH-P.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Imaginal disks;
RC MEDLINE=92146957; PubMed=1346609;
RA Decanilis M., Cheng N.S., Pierre D., Brock H.W.;
RT "The polyhomeotic gene of Drosophila encodes a chromatin protein that
RL shares polyene chromosome-binding sites with Polycomb.";
RN Gene Dev. 6:223-232(1992).
[2]
RP SEQUENCE OF 199-1584 FROM N.A.
RC MEDLINE=92039031; PubMed=1937015;
RA Deatrick J., Daly M., Randsholt N.B., Brock H.W.;
RT "The complex genetic locus polyhomeotic in Drosophila melanogaster
RL potentially encodes two homologous zinc-finger proteins.";
RN Gene 105:185-195(1991).
CC -1- FUNCTION: Binds to polyene chromosome. Seems to interact with
CC Pc. May interact with proteins already bound to promoter
CC complexes and may be a negative regulator of homeotic and
CC segmentation genes. Plays a role in regulating the expression of
CC other pair-rule genes such as eve, ftz, and H.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Salivary glands.
CC -1- SIMILARITY: TO MOUSE EARLY DEVELOPMENT REGULATOR PROTEIN RAE-28.
CC -1- CAUTION: It is uncertain whether Met-1 or Met-9 is the initiator.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; X63672; CAA45211.1; -.
DR EMBL; M64750; -; NOT_ANNOTATED_CDS.
DR PDB; 1KW4; 05-JUN-02.
DR Flybase; Fgn0004861; ph-P.
DR GO; GO:0016458; P:gene allencing; IGI.
DR InterPro; IPR001660; SAM.
DR Pfam; PF00536; SAM; 1.
DR SMART; SM00454; SAM; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.
KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein;
KW 3D-structure.
FT ZN FING. 1365 1387 C4-TYPE.
FT DOMAIN 1513 1577 SAM.
FT DOMAIN 74 80 POLY-GLN.
FT DOMAIN 411 450 GLN-RICH.
FT DOMAIN 494 520 GLN-RICH.
FT DOMAIN 619 650 GLN-RICH.
FT DOMAIN 775 960 GLN-RICH.
FT DOMAIN 1233 1290 SER/THR-RICH.
FT CONFLICT 254 254 MISSING (IN REF. 2).
FT CONFLICT 1415 1415 D -> A (IN REF. 2).
SQ SEQUENCE 1589 AA; 167297 MW; A6DFCF9106E1891 CRC64;

Query Match 4.8%; Score 133.5; DB 1; Length 1589;
Best Local Similarity 24.6%; Pred. No. 0.67;
Matches 86; Conservative 43; Mismatches 151; Indels 69; Gaps 15;

QY 214 ANQHRALVEGSAFATVAVSSGSCWDSQAPILLCYLOSFWTGSFILLNPSSRSKQANTL 273
DB 982 ALQALASAGAIFOIAPKPTGCTSSSPSTSVITNQS-----STPLVTSSTVASIQOAOQTQ 1037
QY 274 LGS1HFD---PEAACDDSTFQPCSPRALANKYVDPSRSIYTLNDG--LSQSEAVAVGR 329
DB 1038 SAQVHOHQQLISATLTAAGTQOQPGPSSLPTTNPLTAMTSMNATVGHLSLTAIPVTVS- 1096
QY 330 YEDTYTNGNPMFLCTLAAEQLYDALYQWKGSLVVD-----VSLDFPK 376
DB 1097 -YTAIVATSSPOLVILSTASS-----GGGGSIPATPTKETPSPKATATVPIGSPK 1147
QY 377 ALYS--DAATGYSSSSSYSSIVDAVKTFADGVSVETHAASNGSMSEQYKSDGEQL 434
DB 1148 TPVSGKDTCTTPKSPATVAVSVEASSSTGE-----ALSNG-----DASD----- 1188
QY 435 SARDLTWGYAALLTANNRNSV--PASWGETSASVPG-----TCAI-TSAICTYSS 484
DB 1189 --RSSTLKGAATTPYSKQNAAVQPPS--TTPNSVSGKEPKLATCGSLTISATSTSTT 1243
QY 485 VTYTSPSIVATGTTTATPTPGSGSVTSTKTATASKTSTTRSGMS 533
DB 1244 TITTINGIVART--TASTAVSTASTTTTSSGTFITSTCTTTTSSIS 1290

RESULT 41
XANP_XANS2 STANDARD; PRT; 827 AA.
ID XANP_XANS2 STANDARD; PRT; 827 AA.
AC 060106;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Xanthomonas precursor (EC 3.4.21.101) (Xanthomonas aspartic
DE proteinase) (Xanthomonaspepsin) (Carboxyl proteinase) (XCP).
OS Xanthomonas sp. (tetra T-22).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=136420;
RN
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC MEDLINE=97058302; PubMed=8902622;
RX Oda K., Ito M., Uchida K., Shibano Y., Fukushima K.-I., Takahashi S.;
RT "Cloning and expression of an isovaleryl peptidyl-in-sensitive
RT carboxyl proteinase gene from Xanthomonas sp. T-22.";
RL J. Biochem. 120:564-572(1996).

```

```

RN [2]
RP MUTAGENESIS OF ASPARTIC ACID AND GLUTAMIC ACID RESIDUES, AND ACTIVE
RX SITES.
RA MEDLINE=99419069; PubMed=10488127;
RL Oyama H., Abe S.-I., Ushiyama S., Takahashi S., Oda K.;
RT "Identification of catalytic residues of pepstatin-insensitive
RT carboxyl proteinases from prokaryotes by site-directed mutagenesis.";
RL J. Biol. Chem. 274:27815-27822(1999).
CC - CATALYTIC ACTIVITY: Cleavage of casein.
CC - COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC - SUBCELLULAR LOCATION: Secreted.
CC - SIMILARITY: Belongs to peptidase family S53.
CC - SIMILARITY: Contains 1 PKD domain.
CC -
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D83740, BAI12093.1, -.
DR MEROPS: S53.002; -.
DR InterPro: IPR000601, PKD.
DR InterPro: IPR007280, PPC.
DR Pfam: PF00082; Peptidase_S8, 1.
DR Pfam: PF00801; PKD, 1.
DR Pfam: PF04151; PPC, 1.
DR SMART: SM00089; PKD, 1.
DR PROSITE: PS50093; PKD, 1.
DR HydroLase: Protease; Serine protease; Zymogen; Signal;
KM Calcium-binding.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 237 REMOVED IN MATURE FORM.
FT CHAIN 238 635 XANTHOMONALISIN.
FT PROPEP 636 827 REMOVED IN MATURE FORM.
FT DOMAIN 635 722 PKD.
FT ACT_SITE 312 312 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 316 316 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 544 544 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT METAL 585 585 CALCIUM (VIA CARBONYL OXYGEN)
FT METAL 586 586 CALCIUM (VIA CARBONYL OXYGEN)
FT METAL 603 603 CALCIUM (VIA CARBONYL OXYGEN)
FT METAL 603 603 CALCIUM (BY SIMILARITY).
FT METAL 605 605 CALCIUM (BY SIMILARITY).
SQ SEQUENCE 827 AA; 83706 MW; 21A33C4C683DBC8F CRC64;
Query Match 4.8%; Score 132; DB 1; Length 827;
Best Local Similarity 19.9%; Pred. No. 0.35;
Matches 122; Conservative 72; Mismatches 221; Indels 197; Gaps 27;
QY 16 GLANVISKRATLDSWLSNEATV-----ARVAILNNGIGAGWVGADSGIVVAPST 67
DB 205 GLQNVSVKHLHRYHDEHYVPGPNVGTQAAVAHHPDFAIYGGSS-----LPAAT 260
QY 68 DNPDPYFTWTRDSGLVUKTVLVL--FKNQDLSLSTENISAAIYQGISNPSGDLSSG 125
DB 261 NTAVGIIITW---GSTQYTVLDSFTSG--AGLATVNSSTIT-----KVGSG 301
QY 126 AGLEPKFNVDETAFTYSGWGRPORD--GPALRATAMIGF-----GQWLLDNG---Y 171
DB 302 TFRANDPDSN-----GEMSLDSQDIVGIAGVYQQLIFYISANGDSSSGITTDGITASY 354
QY 172 TSTATDIWPLVRNDLSVYAQYNQGYDLMEEVNGSSFETIAVQHRALVEGSAFATAVG 231
DB 355 NRAYTDNIAXKLINVLGDEDTAAQSGTQAADD-----ALFQGAVAAGQGFSTIASG 405
QY 232 SSGCWCSQAPEILCYIQSFWTGSIILANPSSSGSDANTLLGSHTFPEAACDSTF 291
DB 406 -----DAGYQWST---DFTSG----- 419

```

```

QY 292 QPCSPALANHKKEVDSFHSIYTLNDGLSDSEAVAGRRYPEDTYNGNPWF-----L 343
DB 420 ---SPGVANNSAGTVKIDLTHYSSEPPASPYIYQVGCTTLST--SGTTWSGETVWNEGL 474
QY 344 CTLAABE-----QLYDALYQMDKQSGLEVTVDLSLDF--FKLLYSPAATGYTS 388
DB 475 SAIAFQGGDNNOFLMATGCGVSLYEAPSPQSSVSSSTRKVGPDPLAFDAASSGALLIVN 534
QY 389 SSSSTSSSIYDAKTPADGFVSIVETHAASNGSM---SEQYDKSDGEQLSARDLTWSYA 444
DB 535 GSTEQYQGTSLASPLVGAFARI---ESAANNAIGPASKFYQAFPTQSLIHDVT----- 587
QY 445 ALLITANN-----RRNSVPASW-----GETSASSVPG-T 472
DB 588 ---SGNNGYQSHGYTATGFEATGFGSPDICKLNTVYAQMNVYTGCGGSGTNAAPVANSF 644
QY 473 CAATSAIGTYSSTVYVSWPQIVV-----TGCITTTATPT-----GSGSV--TSNKT 518
DB 645 VATTGVAATFTDSDSDGSIASHAWTFDGGSTVATSPBHTYSAAGTYVAETVTDNAG 704
QY 519 ATASKTSTTTRS 530
DB 705 ATSTKTSVTVS 716

RESULT 42
FLO1 YEAST STANDARD; PRT; 1537 AA.
ID FLO1 YEAST
AC P32766;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Flocculation protein FLO1 precursor (Flocculin 1).
OS FLO1 OR YAR050W.
ON Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94262325; PubMed=8203162;
RA Watari U., Takata Y., Ogawa M., Sahara H., Koshino S., Omela M.-L.,
RA Aitakinen U., Jaatinen R., Penttilae M., Keranen S.;
RT "Molecular cloning and analysis of the yeast flocculation gene FLO1.";
RL Yeast 10:211-225(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae.";
RT Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RN [3]
RP PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=93289821; PubMed=8511970;
RA Teunissen A.W.R.H., Holub E., van der Hucht J., van den Berg J.A.,
RA Steensma H.Y.;
RT "Sequence of the open reading frame of the FLO1 gene from
RT Saccharomyces cerevisiae.";
RL Yeast 9:423-427(1993).
RN [4]
RP REVIEW.
RX MEDLINE=96076625; PubMed=7502576;
RA Teunissen A.W., Steensma H.Y.;
RT "Review: the dominant flocculation genes of Saccharomyces cerevisiae
RT constitute a new subtelomeric gene family.";
RL Yeast 11:1001-1013(1995).
CC - FUNCTION: May be directly involved in the flocculation process.
CC - SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor

```

(Potential).
 CC -1- PFM: Extensively O-glycosylated (Probable).
 CC -1- SIMILARITY: Belongs to the flocculin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: X78160; CAAS5024.1; -.
 DR EMBL: L28920; AAC09499.1; ALT_SEQ.
 DR PIR: S53465; S53465.
 DR Germonline: 138425; -.
 DR SGD: S0000084; FLO1.
 DR GO: GO:0009277; C:cell wall (sensu Fungi); IDA.
 DR GO: GO:0005194; F:cell adhesion molecule activity; IDA.
 DR GO: GO:000128; P:flocculation; IMP.
 DR InterPro: IPR001389; Flocculin.
 DR Pfam: PF00624; Flocculin; 18.
 KM Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
 FT SIGNAL 1 24
 FT CHAIN 1 24
 FT PROPEP 25 1537
 FT DOMAIN 274 1240
 FT REPEAT 278 1087
 FT REPEAT 278 322
 FT REPEAT 323 367
 FT REPEAT 368 412
 FT REPEAT 413 457
 FT REPEAT 458 502
 FT REPEAT 503 547
 FT REPEAT 548 592
 FT REPEAT 593 637
 FT REPEAT 638 682
 FT REPEAT 683 727
 FT REPEAT 728 772
 FT REPEAT 773 817
 FT REPEAT 818 862
 FT REPEAT 863 907
 FT REPEAT 908 952
 FT REPEAT 953 997
 FT REPEAT 998 1042
 FT REPEAT 1043 1087
 FT DOMAIN 1118 1157
 FT REPEAT 1118 1137
 FT REPEAT 1138 1157
 FT DOMAIN 1226 1392
 FT REPEAT 1226 1276
 FT REPEAT 1291 1341
 FT REPEAT 1342 1392
 FT DOMAIN 1408 1434
 FT REPEAT 1408 1416
 FT REPEAT 1417 1425
 FT REPEAT 1426 1434
 FT REPEAT 1435 1435
 FT CARBOHYD 187 187
 FT CARBOHYD 262 262
 FT CARBOHYD 329 329
 FT CARBOHYD 374 374
 FT CARBOHYD 419 419
 FT CARBOHYD 509 509
 FT CARBOHYD 554 554
 FT CARBOHYD 599 599
 FT CARBOHYD 644 644
 FT CARBOHYD 689 689
 FT CARBOHYD 734 734
 FT CARBOHYD 1114 1114
 FT CARBOHYD 1537 1537
 FT SEQUENCE 1537 AA; 160722 MW; 992650C6B9ABCEA CRC64;
 Query Match 4.7%; Score 130.5; DB 1; Length 1537;

Best Local Similarity 20.0%; Pred. No. 1;
 Matches 119; Conservative 97; Mismatches 236; Indels 143; Gaps 27;
 QY 11 GLVCTGLANVSKRALDLWSLNEATVARTALINMGAGAVSGADSGIVASPESTDNP 70
 DB 769 GLIST-----TTEPW-TGTFSTSTEMTTVTGNGQPTD--ETVIVRTFSEG- 814
 QY 71 DRYTYWTRD--SGLVKTVLDLPRNGDTSLSTIENVI-----SAOAVQGISNP--SGDL 122
 DB 815 --LVTTTPEWGTFTFTSTEMTTTGTGNGVPRDEVIIVIRFTSGLISTTEPEWTGR 872
 QY 123 SSGAGLGEKPFVDEFTAYTGSWGRPORD-----GPALRATAMIGGQWLLDNGYTSTAT 176
 DB 873 TS-----TSMTMTTGTGNGQPTDEVIIVIRFTSGLISTTEPEW--TGTFSTST 922
 QY 177 D-----IWPVANDLSYVAQYNNQGYDMEV-----NGS-SFF 211
 DB 923 EMTVTGTGNGVPRDEFTVIRFTPSGLISTTEPEWTGTFSTSTVTITGTNGQPTDE 982
 QY 212 TIAVGRALVEGSAFATAVGSSCWCDQAPELICVYLGSPWFGSFLANPD-----SSRS 267
 DB 983 TVIVIRTPISGLISTTT-----EPMGTFTSTEMTTVTGNG 1022
 QY 268 KQANTLLGSIHPTPEACDSTFQPCSPRALANHKRVDSFRSIYTLNDGLSDSAVAV 327
 DB 1023 QPTDEVIIVIRFTPSGLVTTT--EPMGTFTSTEM-----STVTGNGVPRDEVIIV 1076
 QY 328 GRYP-----EDTYNGNPNPFLCTLAABEQYDALYMDKQGLSEVTDVSLDFKALYSDA 382
 DB 1077 VKPTTVAISSLSSSSSQITSSITSSRPITTFPYSNGTSVSSSVSSVTSLSFTSS 1136
 QY 383 ---ATGTSSSSSTYSIV-DAVKTFADGFSIVETHAASNGSMGQYKSDG----- 431
 DB 1137 PVTSSVSSSTTSTSTISSESKS-----SVTPSSSTSGS--SSSTSSAGSVSSSR 1189
 QY 432 -EQLSARDLTWSYALTLTANNRNSVVPASGFTSASVPGTCAATSAIGTSSVTVTSW 490
 DB 1190 ISSESSKSPVTSSSLPLVTSATTS-----QETASLSPATTTKTSSEQTLL--VVTSC 1241
 QY 491 PSIVATGTT---TTATPTGSGSVT-----STSKTAAASKTSTTR 529
 DB 1242 ESHVCTESISPAIVSATVTVSGVTEYTWCPISYETTKQTKGTETETTK 1296
 RESULT 43
 ID TENA_PIG STANDARD; PRT; 1746 AA.
 AC Q29116; P98142;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Tenaasin precursor (TN) (Hexabrachion) (Cytocactin) (Neuronectin)
 DE (GMEM) (J1) (Mitochondrial antigen) (Glioma-associated-extracellular
 DE matrix antigen) (GP 150-225) (Tenaasin-C) (TN-C) (P230).
 GN TNC OR HXB.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OC NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS MAJOR; MINOR-1 AND MINOR-2).
 RC TISSUE=Submaxillary gland;
 RX MEDLINE=92104189; PubMed=1722152;
 RA Nishi T., Weinstein J., Gillespie W.M., Paulson J.C.;
 RT "Complete primary structure of porcine tenascin: detection of
 RL tenascin transcript in adult submaxillary glands.";
 RL Eur. J. Biochem. 202;643-648 (1991).
 RN [2]
 RP SEQUENCE OF 813-825; 887-917; 998-1011; 1597-1608 AND 1719-1730.
 RC TISSUE=Fetal brain;
 RX MEDLINE=98158123; PubMed=9498558;
 RA Wakatsuki S., Ho S.H., Arioka M., Yamasaki M., Kitamoto K.;
 RT "Isolation and characterization of a 230 kDa protein (p230)

	FT	DOMAIN	343	373	EGF-LIKE 7.
specifically expressed in fetal brains: its involvement in neurite outgrowth from rat cerebral cortex neurons grown on monolayer of astrocytes.";	FT	DOMAIN	374	404	EGF-LIKE 8.
J. Biochem. 122:1146-1152(1997).	FT	DOMAIN	405	435	EGF-LIKE 9.
-I- FUNCTION: SAM (substrate-adhesion molecule) that appears to inhibit cell migration. May play a role in supporting the growth of epithelial tumors. Is a ligand for integrins alpha-8/beta-1,	FT	DOMAIN	436	466	EGF-LIKE 10.
alpha-9/beta-1, alpha-v/beta-3 and alpha-v/beta-6.	FT	DOMAIN	467	497	EGF-LIKE 11.
-I- FUNCTION: Plays a role during early brain development particularly in growth cone guidance. Involved in neurite outgrowth from cortical neurons grown on the monolayer of astrocytes.	FT	DOMAIN	498	528	EGF-LIKE 12.
-I- SUBUNIT: Hexameric. A homotrimer may be formed in the triple coiled-coil region and may be stabilized by disulfide rings at both ends. Two of such half-hexabrachions may be disulfide linked within the central globule.	FT	DOMAIN	529	559	EGF-LIKE 13.
-I- ALTERNATIVE PRODUCTS:	FT	DOMAIN	560	589	EGF-LIKE 14.
Event-Alternative splicing; Named isoforms=3;	FT	DOMAIN	590	620	EGF-LIKE 15.
Comment=Isoforms are produced in a tissue-and time-specific manner during development;	FT	DOMAIN	621	710	FIBRONECTIN TYPE-III 1.
Name=Minor-2;	FT	DOMAIN	711	801	FIBRONECTIN TYPE-III 2.
IsoId=Q29116-1; Sequence=Displayed;	FT	DOMAIN	802	891	FIBRONECTIN TYPE-III 3.
Name=Major;	FT	DOMAIN	892	983	FIBRONECTIN TYPE-III 4.
IsoId=Q29116-2; Sequence=VSP_001416;	FT	DOMAIN	984	1071	FIBRONECTIN TYPE-III 5.
Name=Minor-1;	FT	DOMAIN	1072	1162	FIBRONECTIN TYPE-III 6.
IsoId=Q29116-3; Sequence=VSP_001417;	FT	DOMAIN	1163	1253	FIBRONECTIN TYPE-III 7.
-I- TISSUE SPECIFICITY: Submaxillary glands and brain.	FT	DOMAIN	1254	1342	FIBRONECTIN TYPE-III 8.
-I- DEVELOPMENTAL STAGE: Predominantly expressed in the embryonic and early postnatal stages. Little or no detection in adult brain.	FT	DOMAIN	1343	1430	FIBRONECTIN TYPE-III 9.
-I- INDUCTION: By TGF-beta.	FT	DOMAIN	1431	1518	FIBRONECTIN TYPE-III 10.
-I- SIMILARITY: Contains 15 EGF-like domains.	FT	DOMAIN	1527	1733	FIBRINOGEN C-TERMINAL.
-I- SIMILARITY: Contains 1 fibronectin type III domain.	FT	DISULFID	64	64	INTERCHAIN (POTENTIAL).
This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announce/or send an email to license@ebi.ac.uk).	CC				
-----	CC				
EMBL; X61599; GAA43796.1; --	DR	EMB1; X61599; GAA43796.1; --	FT	DISULFID	381
PIR; S19694; S19694.	DR	HSSP; P24821; ITEN.	FT	DISULFID	394
InterPro: IPR0060209; EGF_lke.	DR	InterPro: IPR0060209; EGF_lke.	FT	DISULFID	403
InterPro: IPR008957; FN_III-like.	DR	InterPro: IPR008957; FN_III-like.	FT	DISULFID	418
InterPro: IPR003961; FN_III.	DR	InterPro: IPR003961; FN_III.	FT	DISULFID	423
InterPro: IPR006210; IRGF.	DR	InterPro: IPR006210; IRGF.	FT	DISULFID	439
InterPro: IPR002049; Laminin_EGF.	DR	InterPro: IPR002049; Laminin_EGF.	FT	DISULFID	443
Pfam: PF00008; EGF_12.	DR	Pfam: PF00008; EGF_12.	FT	DISULFID	454
Pfam: PF00147; fibrinogen_C_1.	DR	Pfam: PF00147; fibrinogen_C_1.	FT	DISULFID	465
Pfam: PF00044; fn3; 10.	DR	Pfam: PF00044; fn3; 10.	FT	DISULFID	480
PRINTS; PR00011; EGFLAMININ.	DR	PRINTS; PR00011; EGFLAMININ.	FT	DISULFID	487
SMART; SMO0181; EGF_8.	DR	SMART; SMO0181; EGF_8.	FT	DISULFID	496
SMART; SMO0186; FBG; 1.	DR	SMART; SMO0186; FBG; 1.	FT	DISULFID	501
SMART; SMO0060; FN3; 8.	DR	SMART; SMO0060; FN3; 8.	FT	DISULFID	511
PROSITE; PS00022; EGF_1; 15.	DR	PROSITE; PS00022; EGF_1; 15.	FT	DISULFID	516
PROSITE; PS01186; EGF_2; 14.	DR	PROSITE; PS01186; EGF_2; 14.	FT	DISULFID	527
PROSITE; PS50026; EGF_3; 5.	DR	PROSITE; PS50026; EGF_3; 5.	FT	DISULFID	532
Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;	KW	Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;	FT	DISULFID	547
Extracellular matrix; Alternative splicing; Signal.	KW	Extracellular matrix; Alternative splicing; Signal.	FT	DISULFID	558
SIGNAL.	FT	POTENTIAL.	FT	CARBOHYD	563
CHAIN	FT	TENASCIN.	FT	CARBOHYD	578
DOMAIN	FT	INVOLVED IN HEXAMER FORMATION.	FT	CARBOHYD	589
DOMAIN	FT	COILED COIL (POTENTIAL).	FT	CARBOHYD	594
DOMAIN	FT	EGF-LIKE 1 (INCOMPLETE).	FT	CARBOHYD	609
DOMAIN	FT	EGF-LIKE 2.	FT	CARBOHYD	611
DOMAIN	FT	EGF-LIKE 3.	FT	CARBOHYD	620
DOMAIN	FT	EGF-LIKE 4.	FT	CARBOHYD	638
DOMAIN	FT	EGF-LIKE 5.	FT	CARBOHYD	640
DOMAIN	FT	EGF-LIKE 6.	FT	VARSPLIC	641
DOMAIN	FT	Missing (in isoform Major).	FT		

FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 706 AA; 75455 MW; 5E7BF2C77780DC2 CRC64;
 Query Match 4.7%; Score 130; DB 1; Length 706;
 Best Local Similarity 17.6%; Pred. No. 0.39; Indels 282; Gaps 35;
 Matches 130; Conservative 111; Mismatches 217;
 QY 1 MSFRSLALSGVCTGLANYI-SKRATLDSML-----SNEATVARTAILNIGADGAW 52
 DB 1 MQRNLILQASSLI-SGLSLADSSSTTGCGVAPSIIPCPDDSTLVANA----- 48
 QY 53 VSGADSGIVVASTNDPDTFYTRSGVLTKLVLPFN-----GDTSLSTIENYI 106
 DB 49 -----SGLSTA-ETD-----WLKKRDAYTKKALHSFLSRATSNPDSLTSLTFSSN 94
 QY 107 SAQAVIGGINPSGD---LSSGAGLGPKEFVDEATATGSGRQGRPALRYATMGF- 162
 DB 95 SSNVPKIGLACSGGCTYAMLGAGMTIAMDNRIDGANEHLG-----GLIQSSTYLSGLS 149
 QY 163 -GQMLD-----NGYTS-----TATDIWPLVR-----NDLSYVAQYW----- 194
 DB 150 GGMWLTGTLMNMTSTVOELVDHMSDSIMNITKSIYVPGSGSLVTITRMBSIVGEVQ 209
 QY 195 --NOTG-----DIMEVNGSSFP----- 211
 DB 210 AKSDAGNISLSDLMARALSYNFPPLPDAGSALTWSLRDVPFKNGEMPLITVADGR 269
 QY 212 ----TIAVORALVEGSAFATAVSSQWCDSQAPELICVLIQSPWTSFTL---ANFDS 263
 DB 270 YPGTTVINLNAATLFEPTFEMG-----SW-----DPSLNAFTDVKYLQITVNTNGKPVKQ 320
 QY 264 SRSQKD-ANTILGSIHTFDEPAACDDSTFQPCSEPRALANKEVDSFRSIYTLNDGLSDS 322
 DB 321 CVSGDYAGRVIAVTSASLNFNEFSLDST-----STYYKMINSPANKY-VNNLSQDD 370
 QY 323 EAVAVGHPEDTYNGPW-----FLCTLAALAAQLY----- 353
 DB 371 DDIAI-----YANPFXDTEFVDNRNYSIYDADDFLVGDGEDQNLPLVPLIKKE 422
 QY 354 ---DALY-----OMDKOG-----SLEVTD- 369
 DB 423 RDDDVVALIISDNTDSWPSGCVMTNTYERQYSKQKGMAFPVPDPVNTPLNGLTKNP 482
 QY 370 -----VSLDFPKAL---YSDAATGYSSSSSTYS---SIVDAKTPADGFSIVET 414
 DB 483 TFFGCAKNTLDELYIPPLVYIINTGHSFNGSTLKNMYNTERIGMIRNG----- 536
 QY 415 HAASNGMSSEYD-----KSDGEQLSA-----RDLTSTYALALTANNRN 454
 DB 537 EAATMGFTDSDNFCIGICGAIIRKQESINATLPRPCTKCFADYCNWGTLSSTAN--- 592
 QY 455 SVVPASGSETASVPGTCAATSAIG-TYSSVYTWSPSIVANGGTTTATPGSGSVTS 513
 DB 593 -----PELSGNSSTYSGAIAHSISATDGPITALLGSTGNTSTSTSSNVT 645
 QY 514 TSKTTATASKTSTTTRSGMS 533
 DB 646 NSNNSSTTLNNSSSSSIS 665
 RESULT 45
 AMYH YEAST STANDARD; PRT, 1367 AA.
 ID AMYH YEAST P08068; P08068;
 AC P08068; P08068; Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 31, Last sequence update)
 DT 10-FEB-1995 (Rel. 42, Last annotation update)
 DE Glucoamylase (SI/S2 precursor (EC 3.2.1.3) (Glycan 1,4-alpha-
 GN glucosidase) (1,4-alpha-D-glucan glucosylase).
 OS STAL OR STAL OR MAL5 OR YIR019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Bukatayota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 ON NCBI TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RX MEDLINE=97313266; PubMed=9169870;
 RA Chutcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
 RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
 RA Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
 RA Moulé S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
 RA Skelton J., Smith V., Walsh S., Whitehead S., Barrett B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX";
 RL Nature 387:84-87(1997).
 RN [2]
 RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
 RX MEDLINE=87194600; PubMed=3106330;
 RA Yamashita I., Nakamura M., Fukui S.;
 RT "gene fusion is a possible mechanism underlying the evolution of
 STAL.";
 RL J. Bacteriol. 169:2142-2149(1987).
 RN [3]
 RP SEQUENCE OF 1-31 FROM N.A.
 RC STRAIN=SPX101-1C;
 RX MEDLINE=89031230; PubMed=3141213;
 RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
 RT "Similar short elements in the 5' regions of the STA2 and SGA genes
 from Saccharomyces cerevisiae.";
 RL FEBS Lett. 239:179-184(1988).
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -I- SIMILARITY: TO S.POMBE SPEC215.13.
 CC -I- SIMILARITY: SOME, TO S.POMBE SPEC285.13C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z38061; CA86176.1; -;
 DR EMBL: M16164; AA35014.1; -;
 DR EMBL: M16165; AA35015.1; -;
 DR EMBL: X13857; CA32069.1; -;
 DR PIR: S48478; S48478.
 DR Genmonline: 139731; -;
 DR SGD: S0001458; MUC1.
 DR GO: GO:0005886; C:plasma membrane, IDA.
 DR GO: GO:0030447; P:filamentous growth; IDA.
 DR GO: GO:0007125; P:invasive growth; IMP.
 DR GO: GO:0007124; P:pseudophal growth; IMP.
 KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW Signal; Multigene family.
 FT SIGNAL 1 21
 FT CHAIN 1 21
 FT DOMAIN 210 1367
 FT CARBOHYD 817 817
 FT CARBOHYD 874 874
 FT SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;
 Query Match 4.7%; Score 129.5; DB 1; Length 1367;
 Best Local Similarity 20.4%; Pred. No. 1;
 Matches 111; Conservative 76; Mismatches 211; Indels 147; Gaps 22;
 QY 29 SMTSNEATYARTAILNIGADGAWSGADSGIVVASTNDPDTFYTRSGVLTKLV 88
 DB 13 SLTFNSALGPPTLALVPGSSBEGT-----SCNSIVNGCPND-----FNMHNDQONIMQYTL 63
 QY 89 DLFRRNDTSLSTIENYISQAIVOGISNPSGDLSSGAGLGPKEFVND-----ETAY 140
 DB 64 DV-----TSVSWVDNTYQITTHVKG-----KENIDLKLYMSLKIGV 101

```
QY 141 TSGWGRPORDBALRATAMIGFQ--WLDNGYTSTAT-----DI-----VW-PLVRND 186
Db 102 TGBKGVQ-----LYGVNENTYLDNPTDFATPEVYATQDVNSQVMMPNFIQ 151
QY 187 LSTV-----AQY-----WNOTGYDLMEBVN-----GSS-----FTIAVQHRALVEGSA 225
Db 152 FEYLOQSSAAQYASWMOGTTSPDLSTGCNNYDNQHSQTDPPGFYWNIDCDNNC---GGT 208
QY 226 PATAVGSSCSWCSQAPEILCYLOSPWTSFILANPDSSRSRGKDNLTLSIHTPDPEA 285
Db 209 KSSTTTSST-----ESSTTTSSTESSTTTSSTESSTTTS---TSSS 251
QY 286 CDDSTFQPCSPRALANHKEVVDSEFRSITYTLNDGLSDSEAVAAGRYPEDITYNGNPMFLCT 345
Db 252 TSSSTTAPATPTTTSCTKE-----KPTPTTTSCTKEKPTPPHDTTP---CT 296
QY 346 LAAAEQLYDALYQMDXGSLVETDVSIDFPKALYSDATGYSSSSSTYSIYDAVKTPA 405
Db 297 KKTTSKTCCTKK-----TTPPVPTPSSSTESSAPVPPTPSSSTESSAPVTSST 349
QY 406 DGFVSIIVETHAANGSMSEQYDKSDGEQLSARDLTWSYALLTANRRNSVVPASMGETS 465
Db 350 ESSSAPVPPTPSSST-----TESSSAPVTSSTESSSAPVTSSTTE 389
QY 466 ASSVPGTCAATSAIGTVSSVTWSPSIVATGTTTATPTGSGSVTSTSKTTATASKTS 525
Db 390 SSSAPVPTPSSSTTES-SSAPVTSSTTESAPVTSSTTESSSAPVTSSTTESSSAPVTS 448
QY 526 TTTRS 530
Db 449 STTES 453
```

Search completed: June 28, 2004, 07:43:51
Job time : 24 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:41:12 ; Search time 21 Seconds
(without alignments)
2446.013 Million cell updates/sec

Title: US-10-038-723-2
Perfect score: 2771
Sequence: 1 MSFRSLALSLGLVCTGLANV.....SKTTATASKSTTTTSGMGL 534

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 100 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2771	100.0	534	2	A29776
2	2742	99.0	640	1	ALASGR
3	2742	99.0	640	2	A29166
4	2620.5	94.6	639	2	JQ0607
5	2588.5	93.4	639	2	JT0479
6	1935	69.8	612	2	J01346
7	1577	56.9	626	2	S36364
8	1577	56.9	626	2	T49625
9	1497	54.0	493	2	JC6538
10	1321	47.7	616	2	T33908
11	781.5	28.2	450	2	T39433
12	724	26.1	604	1	JP0001
13	676.5	24.4	519	1	A54549
14	597	21.5	549	1	ALBYG
15	522	18.8	778	1	S48474
16	521	18.8	767	1	JU0474
17	221.5	8.0	615	2	A64501
18	181	6.5	1588	2	A86036
19	181	6.5	1588	2	H91188
20	172	6.2	622	2	G90250
21	170.5	6.2	1275	2	T33369
22	166.5	6.0	107	2	B60754
23	161	5.8	1063	2	D86731
24	160	5.8	888	2	T46726
25	159	5.7	1210	2	A25547
26	159	5.7	2271	2	F90073
27	158	5.7	1283	2	T39174
28	157	5.7	1258	2	JQ0188
29	156.5	5.6	2232	2	T34434

30	155	5.6	1322	2	S07053	ice nucleation pro
31	154.5	5.6	1200	1	SNP50	ice nucleation pro
32	153	5.5	1331	2	T41144	hypothetical serin
33	151.5	5.5	1034	2	JC2143	ice nucleation act
34	151.5	5.5	1306	2	S25370	MSB2 protein - yea
35	151.5	5.5	1331	2	A48954	mannan endo-1,4-be
36	150	5.4	1609	2	S25345	probable membrane
37	149	5.4	967	2	S25345	hypothetical prote
38	147	5.3	507	2	S64507	probable membrane
39	146	5.3	612	2	C90419	glucan 1,4 alpha g
40	144.5	5.2	614	2	F86719	hypothetical prote
41	144	5.2	144	2	T11678	hypothetical prote
42	142	5.1	1419	2	T30531	agglutinin-like ad
43	141.5	5.1	644	1	I40712	endo-1,4-beta-xyla
44	141.5	5.1	1441	2	B86807	hypothetical prote
45	140.5	5.1	1567	2	S11672	ice nucleation pro
46	140.5	5.1	1567	2	B95206	cell wall surface
47	140.5	5.1	4776	2	B85547	probable RTX fam1
48	140.5	5.1	5188	2	B85547	hypothetical serin
49	140	5.1	456	2	T38221	HKR1 protein precu
50	139	5.0	1802	2	S69703	hypothetical prote
51	138.5	5.0	402	2	E86185	spore coat protein
52	138.5	5.0	600	2	S07638	hypothetical prote
53	138.5	5.0	1589	2	T13606	hypothetical prote
54	138	5.0	5291	2	F90696	glycoprotein X pre
55	138	5.0	797	1	VGBEX1	membrane glycopro
56	138	5.0	867	2	T45463	alpha-amylase - fl
57	137	4.9	625	2	T41603	probable membrane
58	136	4.9	556	2	S51892	chitinase (EC 3.2.
59	136	4.9	542	2	B47022	membrane glycopro
60	136	4.9	866	2	T45462	hypothetical prote
61	136	4.9	995	2	S50358	hypothetical prote
62	135.5	4.9	1180	2	E86719	cellulose 1,4-beta
63	135.5	4.9	537	1	JU0150	chitinase (EC 3.2.
64	135.5	4.9	660	2	AD0835	large repetitive p
65	134	4.8	1350	2	T30294	invariant surface
66	133.5	4.8	786	2	T16509	hypothetical prote
67	133.5	4.8	1145	2	B75625	hypothetical prote
68	133	4.8	694	2	F92729	TPR-repeat-contain
69	132.5	4.8	520	2	S62521	hypothetical prote
70	132.5	4.8	770	2	T22808	hypothetical prote
71	132	4.8	827	2	JC4900	transferred entry
72	131.5	4.7	856	2	T00349	Avicelase III - As
73	130.5	4.7	1537	2	S53465	flucoculation prote
74	130.5	4.7	1746	1	S19694	tenascin precursor
75	130	4.7	706	2	S53035	probable lysophosp
76	129.5	4.7	534	2	T39903	serine-rich protei
77	129.5	4.7	1367	1	S48478	glucan 1,4-alpha-g
78	129	4.7	1026	2	A48995	paracrySTALLINE su
79	129	4.7	1032	2	T34433	hypothetical prote
80	129	4.7	1035	2	AD3203	autotransporter pr
81	129	4.7	1073	2	C87374	S-layer protein R8
82	128.5	4.6	273	2	T44657	protein GP80 [limpo
83	128.5	4.6	1104	2	S59310	probable membrane
84	128	4.6	1169	2	S38181	flucoculation prote
85	127.5	4.6	776	2	A37035	isoamylase (EC 3.2
86	127.5	4.6	846	2	AD2672	conserved hypotet
87	127.5	4.6	848	2	B97454	hypothetical prote
88	127.5	4.6	881	2	S56032	probable membrane
89	127.5	4.6	1087	2	C84623	cranmembrane olig
90	127.5	4.6	3705	2	AD0123	probable autotrans
91	127	4.6	498	1	FLRC	flagellin - Escher
92	127	4.6	540	2	A47022	chitinase (EC 3.2.
93	127	4.6	1016	2	T41720	hypothetical prote
94	127	4.6	1459	2	T32271	hypothetical prote
95	126.5	4.6	463	2	T38444	hypothetical prote
96	126.5	4.6	776	2	S13470	isoamylase (EC 3.2
97	126.5	4.6	1203	2	S27545	pullulanase - Ther
98	126	4.5	389	2	T33340	hypothetical prote
99	126	4.5	609	2	S62518	hypothetical prote
100	126	4.5	973	2	T40778	hypothetical 129.5

ALIGNMENTS

RESULT 1
A29776
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) G2 precursor - Aspergillus awamori
N;Alternate names: glucamylase G2
C;Species: Aspergillus awamori
C;Date: 20-Jun-1989 #sequence_revision 20-Jun-1989 #text_change 12-Jun-2003
C;Accession: A93066; A29776
R;Number, J.H.; Meade, J.H.; Cole, G.; Lawyer, F.C.; McCabe, P.; Schweickart, V.; Tal,
Mol. Cell. Biol. 4, 2306-2315, 1984
A;Title: Molecular cloning and characterization of the glucamylase gene of Aspergillus
A;Reference number: A93066; MUID:85085934; PMID:6440004
A;Accession: A93066
A;Molecule type: DNA
A;Residues: 1-534 <NN>
A;Cross-references: GB:K02465; NID:9454405; PID:AA859297.1; PID:g166506
R;Number, J.H.; Meade, J.H.; Cole, G.; Lawyer, F.C.; McCabe, P.; Schweickart, V.; Tal,
submitted to Genbank, February 1985
A;Reference number: A94514
A;Contents: annotation; revisions to the DNA sequence and coding regions for G2 form
C;Comment: The DNA sequence was obtained from Genbank, release 55.0.
C;Superfamily: glucan 1,4-alpha-glucosidase, yeast type; glucan 1,4-alpha-glucosidase, h
C;Keywords: alternative splicing; extracellular protein; glycoprotein; glycosidase; hyd
F;23-447/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match 100.0%; Score 2771; DB 2; Length 534;
Best Local Similarity 100.0%; Pred. No. 5,9e-179;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFRLSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
Db 1 MSFRLSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60

Qy 61 VVASPSTNDPDPYFTWTRDSGLVLTVDLFRNGDLSLSTINYSIAQAIVGISINPSG 120
Db 61 VVASPSTNDPDPYFTWTRDSGLVLTVDLFRNGDLSLSTINYSIAQAIVGISINPSG 120

Qy 121 DLSSGAGLGPKNVDETAATGSMGRPORDGAPALRATAMIGFQOMLNDNGYSTATDIW 180
Db 121 DLSSGAGLGPKNVDETAATGSMGRPORDGAPALRATAMIGFQOMLNDNGYSTATDIW 180

Qy 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTIYVQHRALVEGSAFATAVGSSCSMCDQ 240
Db 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTIYVQHRALVEGSAFATAVGSSCSMCDQ 240

Qy 241 APEILCYLOSFWTGSFTLANFDSRSRGKADNTLLGSIHTPDPEAACDDSTFOCPSPALA 300
Db 241 APEILCYLOSFWTGSFTLANFDSRSRGKADNTLLGSIHTPDPEAACDDSTFOCPSPALA 300

Qy 301 NHKEVDSFSSITYTLNDGLSDSEAVAVGRYPEDTYNGNPFCLCTLAABQLYDALYQMD 360
Db 301 NHKEVDSFSSITYTLNDGLSDSEAVAVGRYPEDTYNGNPFCLCTLAABQLYDALYQMD 360

Qy 361 KQSLLEVTVDSLDFPKALYSDAATGTYSSSSSTYSIVDAVKTFADGFVSVETHAASNG 420
Db 361 KQSLLEVTVDSLDFPKALYSDAATGTYSSSSSTYSIVDAVKTFADGFVSVETHAASNG 420

Qy 421 SMSROYKSDGEOLASADLTWVYALLLTANRRNSVVPASMGFTSASVPTCAATSAIG 480
Db 421 SMSROYKSDGEOLASADLTWVYALLLTANRRNSVVPASMGFTSASVPTCAATSAIG 480

Qy 481 TYSSTVTVSWPSIVATGTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 534
Db 481 TYSSTVTVSWPSIVATGTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 534

RESULT 2
ALASGR
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Aspergillus niger
N;Alternate names: 1,4-alpha-D-glucan glucosylhydrolase; glucamylase
C;Species: Aspergillus niger

C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 02-Jun-2003
C;Accession: A90986; A91161; A05287; A22149; A25402
R;Boel, E.; Hansen, M.T.; Hjort, I.; Hoegh, I.; Fall, N.P.
EMBO J. 3, 1581-1585, 1984
A;Title: Two different types of intervening sequences in the glucamylase gene from Asper
A;Reference number: A90986; MUID:84261458; PMID:6204865
A;Accession: A90986
A;Molecule type: DNA
A;Residues: 1-640 <DB>
A;Cross-references: GB:X00712; GB:K02466; NID:92342; PID:CAA25303.1; PID:g2343
A;Note: the authors translated the codon GAT for residue 317 as Asn
R;Stenason, B.; Larsen, K.; Gunnarsson, A.
Eur. J. Biochem. 154, 497-502, 1986
A;Title: Characterization of a glucosylase G2 from Aspergillus niger.
A;Reference number: A91161; MUID:86136085; PMID:3081341
A;Contents: comparison of forms G1 and G2
A;Accession: A91161
A;Molecule type: protein
A;Residues: 25-640 <SV>
C;Comment: The large molecular form G1 is shown.
C;Comment: Smaller molecular forms of the enzyme, G2, arise by proteolytic cleavage(s) of
live towards soluble poly- and oligosaccharides.
C;Genetics:
A;Introns: 72/1; 167/3; 200/1; 412/3
C;Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha-
C;Keywords: alternative splicing; extracellular protein; glycoprotein; glycosidase; hyd
F;1-24/Domain: signal sequence #status Predicted <SIG>
F;23-447/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
F;25-640/Product: glucan 1,4-alpha-glucosidase G1 #status experimental <GG1>
F;25-538/Product: glucan 1,4-alpha-glucosidase G2, long form #status experimental <GG2>
F;25-536/Product: glucan 1,4-alpha-glucosidase G2, short form #status experimental <GG2>
F;135,419/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;465,467,468,477,483,484,492,508,510,513,515,522,525,527,529,530,532/Binding site: carb
F;476,486,488,496,499,500,502,504,506,512,514,517,518,520,524,526,528,531,534,535/Bit

Query Match 99.0%; Score 2742; DB 1; Length 640;
Best Local Similarity 99.6%; Pred. No. 6.8e-177;
Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSFRLSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
Db 1 MSFRLSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60

Qy 61 VVASPSTNDPDPYFTWTRDSGLVLTVDLFRNGDLSLSTINYSIAQAIVGISINPSG 120
Db 61 VVASPSTNDPDPYFTWTRDSGLVLTVDLFRNGDLSLSTINYSIAQAIVGISINPSG 120

Qy 121 DLSSGAGLGPKNVDETAATGSMGRPORDGAPALRATAMIGFQOMLNDNGYSTATDIW 180
Db 121 DLSSGAGLGPKNVDETAATGSMGRPORDGAPALRATAMIGFQOMLNDNGYSTATDIW 180

Qy 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTIYVQHRALVEGSAFATAVGSSCSMCDQ 240
Db 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTIYVQHRALVEGSAFATAVGSSCSMCDQ 240

Qy 241 APEILCYLOSFWTGSFTLANFDSRSRGKADNTLLGSIHTPDPEAACDDSTFOCPSPALA 300
Db 241 APEILCYLOSFWTGSFTLANFDSRSRGKADNTLLGSIHTPDPEAACDDSTFOCPSPALA 300

Qy 301 NHKEVDSFSSITYTLNDGLSDSEAVAVGRYPEDTYNGNPFCLCTLAABQLYDALYQMD 360
Db 301 NHKEVDSFSSITYTLNDGLSDSEAVAVGRYPEDTYNGNPFCLCTLAABQLYDALYQMD 360

Qy 361 KQSLLEVTVDSLDFPKALYSDAATGTYSSSSSTYSIVDAVKTFADGFVSVETHAASNG 420
Db 361 KQSLLEVTVDSLDFPKALYSDAATGTYSSSSSTYSIVDAVKTFADGFVSVETHAASNG 420

Qy 421 SMSROYKSDGEOLASADLTWVYALLLTANRRNSVVPASMGFTSASVPTCAATSAIG 480
Db 421 SMSROYKSDGEOLASADLTWVYALLLTANRRNSVVPASMGFTSASVPTCAATSAIG 480

Qy 481 TYSSTVTVSWPSIVATGTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 530
Db 481 TYSSTVTVSWPSIVATGTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 530

Db 481 TVSSVTWSPSIVATGCTTTATPTGSGSVTSTKTTATASXTSTSS 530

RESULT 3

A29166

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Aspergillus awamori

N/Alternate names: 1,4-alpha-glucan glucohydrolase precursor; glucoamylase precursor

C/Species: Aspergillus awamori

C/Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 02-Jun-2003

C/Accession: A29166

R/Number: U.H.; Meade, J.H.; Cole, G.; Lawyer, F.C.; McCabe, P.; Schweickart, V.; Tal, M.; Cell. Biol. 4, 2306-2315, 1984

A/Title: Molecular cloning and characterization of the glucoamylase gene of Aspergillus

A/Reference number: A93066; PMID:85085934; PMID:6440004

A/Accession: A29166

A/Molecule type: DNA

A/Residues: 1-640 <NUN>

A/Cross-references: GB:K02465; NID:9454405; PIDN:AAB59296.1; PID:g166505

C/Comment: See also PIR:ALASGR.

C/Genetics:

A/Introns: 72/1; 167/3; 200/1; 398/3

C/Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha

C/Keyword: alternate splicing; extracellular protein; glycoprotein; glucosidase; hyd

F:1-16/Domain: signal sequence #status predicted <SIG>

F:19-640/Product: glucan 1,4-alpha-glucosidase #status predicted <MAT>

F:23-447/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match 99.0%; Score 2742; DB 2; Length 640;

Best Local Similarity 99.6%; Pred. No. 6,8e-177; Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNGADGAWVSGADSGI 60

Db 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNGADGAWVSGADSGI 60

QY 61 VVASPSTNDPDIYFYTWRDGLVLTVDLFRNGDTSLSTINYSQAQIVOGISNPSG 120

Db 61 VVASPSTNDPDIYFYTWRDGLVLTVDLFRNGDTSLSTINYSQAQIVOGISNPSG 120

QY 121 DISSGAGLEPKFNVDETATYSGWRPORDGPRALRTAMIGFQWMLDNGYSTARDIYW 180

Db 121 DISSGAGLEPKFNVDETATYSGWRPORDGPRALRTAMIGFQWMLDNGYSTARDIYW 180

QY 121 DISSGAGLEPKFNVDETATYSGWRPORDGPRALRTAMIGFQWMLDNGYSTARDIYW 180

Db 121 DISSGAGLEPKFNVDETATYSGWRPORDGPRALRTAMIGFQWMLDNGYSTARDIYW 180

QY 181 PLVRNDLSVYAQWNOGTGYDLMEVNGSSFFITAVGHRALVBSAFAVAVGSSCSWCDQ 240

Db 181 PLVRNDLSVYAQWNOGTGYDLMEVNGSSFFITAVGHRALVBSAFAVAVGSSCSWCDQ 240

QY 241 ABEILCYLOSFWTGSFTLANFDSRSRSGKANTLLGSIHFPDPAACDDSTFOCSPRALA 300

Db 241 ABEILCYLOSFWTGSFTLANFDSRSRSGKANTLLGSIHFPDPAACDDSTFOCSPRALA 300

QY 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360

Db 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360

QY 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360

Db 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360

QY 361 KQGSLEVTDSLDFFKALYSDAATGYSSSSSTYSISIVAVKTFADGPFYSIVETHAASNG 420

Db 361 KQGSLEVTDSLDFFKALYSDAATGYSSSSSTYSISIVAVKTFADGPFYSIVETHAASNG 420

QY 421 SMSEQYDKSDGELSARDLTWSTYALLLTANRRNSVVPASWGETSASVPGTCAATSAIG 480

Db 421 SMSEQYDKSDGELSARDLTWSTYALLLTANRRNSVVPASWGETSASVPGTCAATSAIG 480

QY 481 TVSSVTWSPSIVATGCTTTATPTGSGSVTSTKTTATASXTSTSS 530

Db 481 TVSSVTWSPSIVATGCTTTATPTGSGSVTSTKTTATASXTSTSS 530

RESULT 4

J00607

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Aspergillus sp.

N/Alternate names: glucoamylase

C/Species: Aspergillus sp.

C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 02-Jun-2003

C/Accession: J00607

R/Shibuya, I.; Gomi, K.; Jimura, Y.; Takahashi, K.; Tamura, G.; Hara, S.

Agric. Biol. Chem. 54, 1905-1914, 1990

A/Title: Molecular cloning of the glucoamylase gene of Aspergillus shiroyamii and its exi

A/Reference number: J00607; PMID:91182400; PMID:1368603

A/Accession: J00607

A/Molecule type: DNA

A/Residues: 1-639 <SHI>

A/Experimental source: strain RIB 2504

C/Comment: This enzyme catalyzes the release of glucose from the non-reducing ends of st

C/Genetics:

A/Introns: 72/1; 166/3; 199/1; 411/3

C/Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha-

C/Keyword: glycosidase; hydrolase; polysaccharide degradation

F:1-24/Domain: signal sequence #status predicted <SIG>

F:23-446/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

F:25-639/Product: glucoamylase #status predicted <MAT>

Query Match 94.6%; Score 2620.5; DB 2; Length 639;

Best Local Similarity 94.3%; Pred. No. 1e-168; Matches 500; Conservative 16; Mismatches 13; Indels 1; Gaps 1;

QY 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNGADGAWVSGADSGI 60

Db 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNGADGAWVSGADSGI 60

QY 61 VVASPSTNDPDIYFYTWRDGLVLTVDLFRNGDTSLSTINYSQAQIVOGISNPSG 120

Db 61 VVASPSTNDPDIYFYTWRDGLVLTVDLFRNGDTSLSTINYSQAQIVOGISNPSG 120

QY 121 DISSGAGLEPKFNVDETATYSGWRPORDGPRALRTAMIGFQWMLDNGYSTARDIYW 180

Db 121 DISSGAGLEPKFNVDETATYSGWRPORDGPRALRTAMIGFQWMLDNGYSTARDIYW 180

QY 181 PLVRNDLSVYAQWNOGTGYDLMEVNGSSFFITAVGHRALVBSAFAVAVGSSCSWCDQ 240

Db 181 PLVRNDLSVYAQWNOGTGYDLMEVNGSSFFITAVGHRALVBSAFAVAVGSSCSWCDQ 240

QY 241 ABEILCYLOSFWTGSFTLANFDSRSRSGKANTLLGSIHFPDPAACDDSTFOCSPRALA 300

Db 241 ABEILCYLOSFWTGSFTLANFDSRSRSGKANTLLGSIHFPDPAACDDSTFOCSPRALA 300

QY 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360

Db 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360

QY 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360

Db 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYQMD 360

QY 361 KQGSLEVTDSLDFFKALYSDAATGYSSSSSTYSISIVAVKTFADGPFYSIVETHAASNG 420

Db 361 KQGSLEVTDSLDFFKALYSDAATGYSSSSSTYSISIVAVKTFADGPFYSIVETHAASNG 420

QY 421 SMSEQYDKSDGELSARDLTWSTYALLLTANRRNSVVPASWGETSASVPGTCAATSAIG 480

Db 421 SMSEQYDKSDGELSARDLTWSTYALLLTANRRNSVVPASWGETSASVPGTCAATSAIG 480

QY 481 TVSSVTWSPSIVATGCTTTATPTGSGSVTSTKTTATASXTSTSS 530

Db 481 TVSSVTWSPSIVATGCTTTATPTGSGSVTSTKTTATASXTSTSS 530

RESULT 5

J00479

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Aspergillus awamori

N/Alternate names: glucoamylase I

C/Species: Aspergillus awamori

C/Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 02-Jun-2003

C/Accession: J00479

R/Hayashida, S.; Kuroda, K.; Ohta, K.; Kuhara, S.; Fukuda, K.; Sakaki, Y.

Agric. Biol. Chem. 53, 923-929, 1989

A/Title: Molecular cloning of the glucoamylase I gene of Aspergillus awamori var. kawachi

A/Reference number: J00479

A/Accession: J00479

A/Molecule type: DNA

A;Residues: 1-639 <HAY>
A;Experimental source: var. kawachi
C;Genetics:
A;Gene: GAI
A;introns: 72/1; 166/3; 199/1; 411/3
C;Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-24/Domains: signal sequence #status predicted <SIG>
F;23-44/Domains: glucan 1,4-alpha-glucosidase homology <GAG>
F;25-639/Product: glucan 1,4-alpha-glucosidase #status predicted <MAT>
F;494-538/Region: raw-starch-affinity region

Query Match 93.4%; Score 2588.5; DB 2; Length 639;
Best Local Similarity 93.8%; Pred. No. 1.5e-166;
Matches 497; Conservative 15; Mismatches 17; Indels 1; Gaps 1;

```
QY 1 MSFRLALSGVCTGIAANVSKRATLDSWLSNEATVARTALINNIGADGAWMSGADSI 60
DB 1 MSFRLALSGVCTGIAANVSKRATLDSWLSNEATVARTALINNIGADGAWMSGADSI 60
QY 61 VVASPTDNDPFFYFTWRDGLVKTIVDLFRNGDTLSLTENYISAQAIVOGINSNP 120
DB 61 VVASPTDNDPFFYFTWRDGLVKTIVDLFRNGDTLSLTENYISAQAIVOGINSNP 120
QY 121 DLSGAGLGEPEKFNVDERTAYTSGWGRPQDGPALRATAMIFGQWLLDNGYSTATDI 180
DB 121 DLSGAGLGEPEKFNVDERTAYTSGWGRPQDGPALRATAMIFGQWLLDNGYSTATDI 180
QY 121 DLSGAGLGEPEKFNVDERTAYTSGWGRPQDGPALRATAMIFGQWLLDNGYSTATDI 180
DB 121 DLSGAGLGEPEKFNVDERTAYTSGWGRPQDGPALRATAMIFGQWLLDNGYSTATDI 180
QY 181 PLVANDLSYVAQVYNQGYDLMEEVNCGSFPTIAVOHRALVEGSAFATAVGSSCWCDC 240
DB 181 PLVANDLSYVAQVYNQGYDLMEEVNCGSFPTIAVOHRALVEGSAFATAVGSSCWCDC 240
QY 180 PLVANDLSYVAQVYNQGYDLMEEVNCGSFPTIAVOHRALVEGSAFATAVGSSCWCDC 239
DB 180 PLVANDLSYVAQVYNQGYDLMEEVNCGSFPTIAVOHRALVEGSAFATAVGSSCWCDC 239
QY 241 APEILCYLGSEFWTGSFILLANFDSRSRSGKDANTLLGSIHTPEPAACDSTFQPCSPRA 300
DB 241 APEILCYLGSEFWTGSFILLANFDSRSRSGKDANTLLGSIHTPEPAACDSTFQPCSPRA 300
QY 240 APEILCYLGSEFWTGSFILLANFDSRSRSGKDANTLLGSIHTPEPAACDSTFQPCSPRA 299
DB 240 APEILCYLGSEFWTGSFILLANFDSRSRSGKDANTLLGSIHTPEPAACDSTFQPCSPRA 299
QY 301 NHKEVDSFRSITVTLNDGLDSEAVAVGRYPEDTYNGNPFCLTLAAEQLYDALYQMD 360
DB 301 NHKEVDSFRSITVTLNDGLDSEAVAVGRYPEDTYNGNPFCLTLAAEQLYDALYQMD 360
QY 300 NHKEVDSFRSITVTLNDGLDSEAVAVGRYPEDTYNGNPFCLTLAAEQLYDALYQMD 359
DB 300 NHKEVDSFRSITVTLNDGLDSEAVAVGRYPEDTYNGNPFCLTLAAEQLYDALYQMD 359
QY 361 KQGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVDAKTPADGFSIVETHAASNG 420
DB 361 KQGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVDAKTPADGFSIVETHAASNG 420
QY 360 KQGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVDAKTPADGFSIVETHAASNG 419
DB 360 KQGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVDAKTPADGFSIVETHAASNG 419
QY 421 SMSQYDKSDGEQSLARDLTWSYALTLTANRRNSVVPASGERTSASSVPGTCAATSAIG 480
DB 421 SMSQYDKSDGEQSLARDLTWSYALTLTANRRNSVVPASGERTSASSVPGTCAATSAIG 480
QY 420 SLSEQPKSDGEQSLARDLTWSYALTLTANRRNSVVPASGERTSASSVPGTCAATSAIG 479
DB 420 SLSEQPKSDGEQSLARDLTWSYALTLTANRRNSVVPASGERTSASSVPGTCAATSAIG 479
QY 481 TYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKSTTTTS 530
DB 481 TYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKSTTTTS 530
QY 480 TYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKSTTTTS 529
DB 480 TYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKSTTTTS 529
```

RESULT 6

QJ01346
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - *Aspergillus oryzae*
N;Alternate names: glucoamylase
C;Species: *Aspergillus oryzae*
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 02-Jun-2003
C;Accession: J01346
R;Hata, Y.; Tsuchiya, K.; Kitamoto, K.; Gomi, K.; Kumagai, C.; Tamura, G.; Hara, S.
Gene 108, 145-150, 1991
A;Title: Nucleotide sequence and expression of the glucoamylase-encoding gene (gluA) frc
A;Reference number: J01346; MUID:92104497; PMID:1761224
A;Accession: J01346
A;Molecule type: DNA
A;Residues: 1-612 <HAT>
A;Cross-references: GB:D10698; DDBJ:D01108; NID:g1160312; PIDN:BA01540.1; PID:g1160313
A;Note: The authors translated the codon TTT for residue 213 as Tyr
C;Comment: This enzyme hydrolyzes starch to glucose.
C;Genetics:
A;Gene: glua
A;introns: 75/1; 169/3; 202/1; 414/3

C;Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;26-449/Domains: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match 69.8%; Score 1935; DB 2; Length 612;
Best Local Similarity 70.3%; Pred. No. 1.4e-122;
Matches 367; Conservative 59; Mismatches 90; Indels 6; Gaps 4;

```
QY 1 MSFRLALSGVCTGIA--NVISKRATLDSWLSNEATVARTALINNIGADGAWMSGADSI 58
DB 2 VSFSSCLRALALGSSVLAQVPLRQATGDTWLSTANRSQAIIINNIGADGQSAQASAP 61
QY 59 GIYVAPSTDNDPFFYFTWRDGLVKTIVDLFRNGDTLSLTENYISAQAIVOGINSNP 118
DB 62 GVIVAPSTDNDPFFYFTWRDGLVKTIVDLFRNGDTLSLTENYISAQAIVOGINSNP 121
QY 119 SGPLSSGAGLGEPEKFNVDERTAYTSGWGRPQDGPALRATAMIFGQWLLDNGYSTATDI 178
DB 122 SGPLSSGAGLGEPEKFNVDERTAYTSGWGRPQDGPALRATAMIFGQWLLDNGYSTATDI 180
QY 179 VWPVANDLSYVAQVYNQGYDLMEEVNCGSFPTIAVOHRALVEGSAFATAVGSSCWCDC 238
DB 181 VWPVANDLSYVAQVYNQGYDLMEEVNCGSFPTIAVOHRALVEGSAFATAVGSSCWCDC 240
QY 239 SQAPELCYLGSEFWTGSFILLANFDSRSRSGKDANTLLGSIHTPEPAACDSTFQPCSPRA 298
DB 241 SQAPVQRCYLGSEFWTGSFILLANFDSRSRSGKDANTLLGSIHTPEPAACDSTFQPCSPRA 300
QY 299 LANHKEVDSFRSITVTLNDGLDSEAVAVGRYPEDTYNGNPFCLTLAAEQLYDALYQ 358
DB 301 LANHKEVDSFRSITVTLNDGLDSEAVAVGRYPEDTYNGNPFCLTLAAEQLYDALYQ 360
QY 359 WDKQGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVDAKTPADGFSIVETHAAS 418
DB 361 WDKQGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVDAKTPADGFSIVETHAAS 420
QY 419 NSGMSQYDKSDGEQSLARDLTWSYALTLTANRRNSVVPASGERTSASSVPGTCAATSA 478
DB 421 TGSMAEQYTKSDGQSLARDLTWSYALTLTANRRNSVVPASGERTSASSVPGTCAATSA 480
QY 479 IGYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKSTTTTS 518
DB 481 IGYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKSTTTTS 521
```

RESULT 7

S36364
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - *Neurospora crassa*
N;Alternate names: glucoamylase; glycoamylase
C;Species: *Neurospora crassa*
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 02-Jun-2003
C;Accession: S36364; S13710; S25539
R;Stone, P.J.; Makoff, A.J.; Parish, J.H.; Radford, A.
Curr. Genet. 24, 205-211, 1993
A;Title: Cloning and sequence analysis of the glucoamylase gene of *Neurospora crassa*.
A;Reference number: S36364; MUID:94037144; PMID:8221928
A;Accession: S36364
A;Molecule type: DNA
A;Residues: 1-626 <SNO>
A;Cross-references: EMBL:X67291
R;Koh-Luar, S.I.; Parish, J.H.; Bleasby, A.J.; Pappin, D.J.C.; Ainley, K.; Johansen, F.B.
Enzyme Microb. Technol. 11, 692-695, 1989
A;Title: Exported proteins of *Neurospora crassa*: 1-glucoamylase.
A;Reference number: S13710
A;Accession: S13710
A;Molecule type: protein
A;Residues: 36-60, 'X', 62, 'X', 64-65 <KOH>
R;Koh-Luar, S.I.; Parish, J.H.; Bleasby, A.J.; Pappin, D.J.C.; Ainley, K.; Johansen, F.B.
submitted to the Protein Sequence Database, January 1990
A;Description: Exported proteins of *Neurospora crassa* 1: - glucoamylase.
A;Reference number: S13711
A;Accession: S13711
A;Molecule type: protein

A:Residues: 36-60,'X',62,'X',64-65 <KOW>
 C:Genetics:
 A:Gene: gla-1
 A:Introns: 82/2
 C:Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha
 C:Keywords: glycosylase; hydrolase; polysaccharide degradation
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-35/Domain: propeptide #status predicted <PRO>
 F:33-458/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
 F:36-626/Product: glucan 1,4-alpha-glucosidase #status experimental <MAT>

Query Match 56.9%; Score 1577; DB 2; Length 626;
 Best Local Similarity 57.6%; Pred. No. 1.9e-98;
 Matches 306; Conservative 85; Mismatches 130; Indels 10; Gaps 6;

```

QY 2 SFRSLALSGVLTGTLANYISKRATLDSWLSNEATYARFALLNNIGADGAWSGADSGIV 61
DB 13 AFQAVLGLDPLDPEKHSHDIKR-SVDSYIQETETPAQKLLCNIGASGRASGVAAGV 71
QY 62 VASPTDNDPYFTWTRDGLVLTVDLPRNG-DTSLSTENTYISQAQIVGINSNSG 120
DB 72 VASPSKSSPDYWTWTRDALVTKLIVDEFTNDYNTLQNTIQAYAAQKLGVSNSPG 131
QY 121 DLSGAGLGEPRKFNVDETAYTSGMGRPQRDGPALRATAMIGFQWMLDNGYTSATDIW 180
DB 132 SLNKGAGLGEPRKFNVDLQFTGAMGRPQRDGPALRATAMIGYKWLVSNGYADTAKSIW 191
QY 181 PLVRNDLSVVAQYWNQYDLMEEVNGSSFTTAVOHRALVEGSAFATAVGSSCMDSQ 240
DB 192 PLYKNDLAVTAQYWNNTGFDLMEEVNSSFTTAAASHRALVEGSAFAKVSQSSCSACDAI 251
QY 241 APEILCYLOSFWTGS-FILANPDSRSRGDANTLGSITFPDEACDDSTPQPCSPRAL 299
DB 252 APQILCFQGSFMSNSGYIISFNVYRSGKDINSVLTSIHFPDPAACDVNTFPCCDRAL 311
QY 300 ANHKEVVDSPRSIYTLNDGLDSEAVAVGRYPEDTYNNGNPMFLCTLAAAEQYDALYOW 359
DB 312 ANHKVVVDSPMR-FWGNVSGRTAGKAAVGRYADVYNGNPMFLATLAAAEQYDALYOW 370
QY 360 DKQGSLEVTDLSDLPFKALYSDAATGYSSSSSTYSIVDAVTFADGFSIVETHAASN 419
DB 371 KKGSLITVSTSLAFKDLVPSVSTGYSSSSSTYTAIINAVTYADGFVDIYAQYTPSD 430
QY 420 GSNSEGYDKSDGEQLSARDLTMSYALLLTANRRNSVVPASMGETSASVPGCAATSAI 479
DB 431 GSIAEFPDQSGAPLSATHLTWSYASFLSAAARRAGIVPWSGAAANSIPGCSASTVA 490
QY 480 GYYSVTVTSWPSIVATGTTTT-ATPTGSGS-----VTSTSKTATASKT 524
DB 491 GSVAATATATSPANLTPASTTTPPTGTGCAADHEVLTFNEKVTTSYGQT 541

```

RESULT 8

T49625
 glucan 1,4-alpha-glucosidase [imported] - Neurospora crassa
 N:Alternate names: protein B5022.70
 C:Species: Neurospora crassa
 C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2003
 C/Accession: T49625
 R/Schulte, U.; Algm, V.; Hohenstein, J.; Brandt, P.; Faermann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, May 2000
 A:Reference number: Z25022
 A:Accession: T49625
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-626 <SCH>
 A:Cross-references: EMBL:ALJ55932; GSPDB:GN00116; NCSP:B5022.70
 A:Experimental source: BAC clone B5022; strain OR74A
 C:Genetics:
 A:Gene: NCSP:B5022.70
 A:Map position: 6
 A:Introns: 82/1
 C:Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha

Query Match 56.9%; Score 1577; DB 2; Length 626;
 Best Local Similarity 57.6%; Pred. No. 1.9e-98;
 Matches 306; Conservative 85; Mismatches 130; Indels 10; Gaps 6;

```

QY 2 SFRSLALSGVLTGTLANYISKRATLDSWLSNEATYARFALLNNIGADGAWSGADSGIV 61
DB 13 AFQAVLGLDPLDPEKHSHDIKR-SVDSYIQETETPAQKLLCNIGASGRASGVAAGV 71
QY 62 VASPTDNDPYFTWTRDGLVLTVDLPRNG-DTSLSTENTYISQAQIVGINSNSG 120
DB 72 VASPSKSSPDYWTWTRDALVTKLIVDEFTNDYNTLQNTIQAYAAQKLGVSNSPG 131
QY 121 DLSGAGLGEPRKFNVDETAYTSGMGRPQRDGPALRATAMIGFQWMLDNGYTSATDIW 180
DB 132 SLNKGAGLGEPRKFNVDLQFTGAMGRPQRDGPALRATAMIGYKWLVSNGYADTAKSIW 191
QY 181 PLVRNDLSVVAQYWNQYDLMEEVNGSSFTTAVOHRALVEGSAFATAVGSSCMDSQ 240
DB 192 PLYKNDLAVTAQYWNNTGFDLMEEVNSSFTTAAASHRALVEGSAFAKVSQSSCSACDAI 251
QY 241 APEILCYLOSFWTGS-FILANPDSRSRGDANTLGSITFPDEACDDSTPQPCSPRAL 299
DB 252 APQILCFQGSFMSNSGYIISFNVYRSGKDINSVLTSIHFPDPAACDVNTFPCCDRAL 311
QY 300 ANHKEVVDSPRSIYTLNDGLDSEAVAVGRYPEDTYNNGNPMFLCTLAAAEQYDALYOW 359
DB 312 ANHKVVVDSPMR-FWGNVSGRTAGKAAVGRYADVYNGNPMFLATLAAAEQYDALYOW 370
QY 360 DKQGSLEVTDLSDLPFKALYSDAATGYSSSSSTYSIVDAVTFADGFSIVETHAASN 419
DB 371 KKGSLITVSTSLAFKDLVPSVSTGYSSSSSTYTAIINAVTYADGFVDIYAQYTPSD 430
QY 420 GSNSEGYDKSDGEQLSARDLTMSYALLLTANRRNSVVPASMGETSASVPGCAATSAI 479
DB 431 GSIAEFPDQSGAPLSATHLTWSYASFLSAAARRAGIVPWSGAAANSIPGCSASTVA 490
QY 480 GYYSVTVTSWPSIVATGTTTT-ATPTGSGS-----VTSTSKTATASKT 524
DB 491 GSVAATATATSPANLTPASTTTPPTGTGCAADHEVLTFNEKVTTSYGQT 541

```

RESULT 9

JC6538
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Aspergillus oryzae
 C:Species: Aspergillus oryzae
 C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 12-Jun-2003
 C/Accession: JC6503; JC6538
 R:Hata, Y.; Ishida, H.; Ichikawa, E.; Kawato, A.; Sugihara, K.; Imai, S.
 Gene 207, 127-134, 1998
 A>Title: Nucleotide sequence of an alternative glucosylase-encoding gene (glaB) express
 A:Reference number: JC6538; MIMD:98172744; PMID:9511753
 A:Accession: JC6538
 A:Molecule type: protein
 A:Residues: 43-53;71-82;136-145;180-187;327-336;393-404 <HAT>
 A:Cross-references: DDBJ:AB07825
 A:Comment: This enzyme is used in industrial processing. It releases glucose from the nor
 C:Genetics:
 A:Introns: 75/1; 344/2
 C:Superfamily: glucan 1,4-alpha-glucosidase, yeast type; glucan 1,4-alpha-glucosidase hon
 C:Keywords: glycoprotein; glycosidase; hydrolase
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:26-450/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
 F:27-493/Product: glucan 1,4-alpha-glucosidase #status experimental <MAT>
 F:139,198,255,369,384,457/Binding site: carbohydrate (Aan) (covalent) #status predicted

Query Match 54.0%; Score 1497; DB 2; Length 493;
 Best Local Similarity 56.9%; Pred. No. 3.3e-93;
 Matches 277; Conservative 80; Mismatches 128; Indels 2; Gaps 2;

QY 5 SLLASGLVCTGLANVISKRAITLDSWLSNEATVARTAIINNIGADGAMVSGADSGIVAS 64
 Db 8 SLNAGAVAHSPFIHKRQSDLNAPFIEAQTPIAKQCYLNNIGADGXLVEGAAGIYVAS 67
 QY 65 PSTNDPVPYFTWTRDSGLVLTUVDLFRNGDTLSLTENYISAQAVQIGINSPSGLS 124
 Db 68 PAKSNPDYFTWTRDAGLTWEEYIEPFIQGDATLESTIQYVVDSDQANBOAVSNPSGSLD 127
 QY 125 GAGLGEPEKFNVDETAYTGSWGRPDGDPALRATAMIGFGQWLDNGYTSATDIWPLVR 184
 Db 128 GSGLAEPKFTYNNISQFLDSWGRPDGDPALRATAMIGFGQWLDNGYTSATDIWPLVR 187
 QY 185 NDLISYVAQYNNQGYDIMEEVNCGSPFTTIAVQHRALVEGSAFATAVGSSCSWCDGQAP 244
 Db 188 NDLISYVAQYNNQGYDIMEEVNCGSPFTTIAVQHRALVEGSAFATAVGSSCSWCDGQAP 246
 QY 245 LCTVLOSFWTGSFLLANDSS-RSGKDANTLIGSIHTFDPEACDDSTFGQCSRALANHK 303
 Db 247 LCHLODFWNGSAVLSNLPITNGRGLDNLISLIGSIHTFDPEACDDSTFGQCSRALANHK 306
 QY 304 EYVDSFPSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTIAAARQLYDALYQMDKOG 363
 Db 307 LVVDSFPSYVGINNGRAGAAAVGPAEDTYQGANWYLTLLVAAELLYDALYQMDKOG 366
 QY 364 SLEVTVDLDFPALYSDAATGYSSSSSTYSSIVDAVKTADGFSIVETHAASNGSMS 423
 Db 367 QVNVTELSLFFPKDLSNVTGSAKSSAYESLTSVATKYADGFIIVQDEYTPDGALA 426
 QY 424 EYVDSFPSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTIAAARQLYDALYQMDKOG 483
 Db 427 EYVDSFPSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTIAAARQLYDALYQMDKOG 486
 QY 484 SVTVYTSW 490
 Db 487 TPTVGSW 493

RESULT 10

S33908
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) P precursor - creosote fungus
 N/Alternate names: glucanase
 C/Species: Amorphytheca resinosa, Hormoconis resinae (creosote fungus)
 C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 02-Jun-2003
 C/Accession: S33908; S36677; S31425; A60754; S31422
 R/Vainio, A.E.I.; Torkkeli, H.T.; Tuusa, T.; Aho, S.A.; Fagerstrom, B.R.; Korhola, M.P.
 Curr. Genet. 24, 38-44, 1993
 A/Title: Cloning and expression of Hormoconis resinae glucanase P cDNA in Saccharomyces
 A/Reference number: S33908; MUID:93365035; PMID:8358830
 A/Molecule type: mRNA
 A/Residues: 1-616 <VAIL>
 A/Cross-references: EMBL:X67708; NID:92785; PIDN:CAA47945.1; PID:92786
 A/Accession: S36677
 A/Molecule type: protein
 A/Residues: 30-106;138-151;168-181;217-223;260-294;310-316;321-334;336-342;369-415;423-4
 R/Joutsen, V.; Torkkeli, T.
 submitted to the EMBL Data Library, August 1992
 A/Reference number: S31425
 A/Accession: S31425
 A/Molecule type: DNA
 A/Residues: 1-616 <JOU>
 A/Cross-references: EMBL:X68143; NID:92787; PIDN:CAA48243.1; PID:92788
 R/Fagerstrom, R.; Vainio, A.; Suoranta, K.; Pakula, T.; Kalkkinen, N.; Torkkeli, H.
 J. Gen. Microbiol. 136, 913-920, 1990
 A/Title: Comparison of two glucanases from Hormoconis resinae.
 A/Reference number: A60754; MUID:90338987; PMID:2116499
 A/Accession: A60754
 A/Molecule type: protein
 A/Residues: 30-53;72-89;145-154;217-223;260-294;423-426; 'X', 428-436 <FAG>
 A/Experimental source: strain ATCC 20495
 C/Genetics:
 A/Genes: gamp

A/Intons: 76/1, 177/3; 347/2
 C/Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha-
 C/Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation
 F/1-29/Domain: signal sequence #status predicted <SIG>
 F/27-455/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
 F/30-616/Product: glucan 1,4-alpha-glucosidase #status experimental <MAT>
 F/99,427/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F/200,514,528,587/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 47.7%; Score 1321; DB 2; Length 616;
 Best Local Similarity 48.6%; Pred. No. 3 2e-81;
 Matches 257; Conservative 88; Mismatches 150; Indels 34; Gaps 6;

QY 14 CTGLANVIS-----KRAITLDSWLSNEATVARTAIINNIGADGAMVSGADSGIVAS 63
 Db 8 CAGALSLTGLSLAIAPTELKARDLSFISERAIALQGLANNIGPDGSAVPGAGAFVVA 67
 QY 64 SPSTNDPVPYFTWTRDSGLVLTUVDLFRNGDTLSLTENYISAQAVQIGINSPSGLS 122
 Db 68 PAKSNPDYFTWTRDAGLTWEEYIEPFIQGDATLESTIQYVVDSDQANBOAVSNPSGSLD 127
 QY 123 SSGAGLGEPEKFNVDETAYTGSWGRPDGDPALRATAMIGFGQWLDNGYTSATDIWPLVR 182
 Db 128 GSGLAEPKFTYNNISQFLDSWGRPDGDPALRATAMIGFGQWLDNGYTSATDIWPLVR 187
 QY 183 VRNDLSYVAQYNNQGYDIMEEVNCGSPFTTIAVQHRALVEGSAFATAVGSSCSWCDGQAP 242
 Db 188 NDLISYVAQYNNQGYDIMEEVNCGSPFTTIAVQHRALVEGSAFATAVGSSCSWCDGQAP 246
 QY 243 EYVDSFPSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTIAAARQLYDALYQMDKOG 300
 Db 247 EYVDSFPSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTIAAARQLYDALYQMDKOG 306
 QY 301 NHKEAVDSFPSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTIAAARQLYDALYQMDKOG 360
 Db 307 NFKVLTDFRNLITTAINGAPREGGVAAGRYAEVYNGNPMFLCTIAAARQLYDALYQMDKOG 366
 QY 367 ARHVLTVDETSIAFFPDYVEVTRVREYKSGNANPQAQIMDAVTAADSVIAAEKYPIS 426
 Db 419 NGSMSEQYKDSDEQISARDLTSYAAALLTANRRNSVPAWGETSASVGTCAATSA 478
 QY 427 NGSLISQFNNDTPTLSAIDLTSYAAFTMSQRRAGQYPSGSRNALPPTTCSASST 486
 QY 479 IGYSSVTVYTSWSPSIVATGGTTATPTGSGVTSKTTATAGTSKSTT 527
 Db 487 PGYI-----TPATAGAPNVTSSCQVSTININATT 517

RESULT 11

T39433
 glucanase precursor - fission yeast (Schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 12-Jun-2003
 C/Accession: T39433
 R/Xiang, Z.; Aves, S.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volkhardt, G.
 submitted to the EMBL Data Library, March 1998
 A/Reference number: Z21854
 A/Accession: T39433
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-450 <LYN>
 A/Cross-references: EMBL:AL022305; PIDN:CAA18423.1; GSPDB:GNO0067; SPDB:SPBC14C8.05C
 A/Experimental source: strain 972h-; cosmid c14C8
 C/Genetics:
 A/Genes: SPDB:SPBC14C8.05C
 A/Map position: 2
 C/Superfamily: glucan 1,4-alpha-glucosidase, yeast type; glucan 1,4-alpha-glucosidase hom

Query Match 28.2%; Score 781.5; DB 2; Length 450;
 Best Local Similarity 37.6%; Pred. No. 4 4e-45;
 Matches 179; Conservative 68; Mismatches 166; Indels 63; Gaps 12;

QY 7 LALSGLVCTG---LANVSKRATLDSEWLSNEATVARTAILNNIGADGAWSGADSGIIVA 63
 Db 8 LILGGVVASBSLSPKRSKEASMDMETOOQKCIANGHMLNNIGDSGMHAKOINPCIIA 67
 QY 64 SPSTNDPDIYFYTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIIVOGISNPSGDL 123
 Db 68 SPSTDPDIYFYQWVRDSALTIMTILDRFEGDKGLEPIIVKYMDEWRLQKVENPSGDFX 127
 QY 124 SGAGIGEPENVDETVTSQMGRRPQRDGRALRATNIGFGOMLJLDGYSTARDIYWPLY 183
 Db 128 AG-GLGEPRFNNVDTSYDGMGRQNDSPALRAIAFLKYNVYLFENGKEVETWIEAV 186
 QY 184 RNDLSVVAQVMNQTDLMEEVNSSFPTIAYOHRALVEGSAFATVAGSSCSWQDAPE 243
 Db 187 LADLDVTAHMTASDLDMEIEDVHYFTLAVQKRAMQDSTAPAKIG-----APD 237
 QY 244 -----ILCYLQSPW-TGSFLLANFDS--SRSGKADANTLGSIHNT--FDPEAACD 287
 Db 238 QALVYRTPEIDILKLGEEFMDPGMGVYKGYKGRVRSGLDSCSTLLASVGENFD----- 291
 QY 288 DSFFQCSRALANHEVVDSPRSITTLNDGLSDSEAVNAGRPEDTYTN-----GNPWF 342
 Db 292 -----WHILEPTLLKDETMTDRDPVNVGMKQ---AMGRYPEVDYDGVSKSIGNPWF 339
 QY 343 LCTLAEOULYDALYOMDKQSLVTDVSLDPFK--ALYSDATATGYSSSSSTYSIYDA 400
 Db 340 ICTSSAAEITIKYAIAYDNKGLEPTEHYNIHFPMKRAFEFD-----PYMWSTIRKN 350
 QY 401 VKTFADGFVSIYETHAASNGSMSEQYDKSDGEOLSARDLTWSYALTLTANRRENV 456
 Db 391 MHTYADNPLKAVAEFQHPNGSMSEQFSRDDGHQKGRDLTWSYSLLNATYREAI 446

RESULT 12

JP0001
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Rhizopus oryzae
 N:Alternate names: 1,4-alpha-D-glucan glucosylhydrolase; glucoamylase
 C:Species: Rhizopus oryzae
 C>Date: 28-Dec-1987 #sequence_revision 23-Aug-1996 #text_change 16-Jun-2000
 C/Accession: JP0001
 R:Ashikari, T.; Nakamura, N.; Tanaka, Y.; Kiuchi, N.; Shibano, Y.; Tanaka, T.; Amachi, T.
 Agric. Biol. Chem. 50, 957-964, 1986
 A>Title: Rhizopus raw-starch-degrading glucoamylase: its cloning and expression in yeast
 A:Reference number: A90022
 A:Accession: JP0001
 A:Molecule type: DNA
 A:Residues: 1-604 <ASH>
 A:Cross-references: GB:D00049; PIDN:BA00033.1; PID:g218035
 A:Experimental source: strain SAM0034
 A>Note: there are two errors in the published sequence (personal communication): GCT (11
 y to the known amino acid sequence of 11 peptides from glucoamylase, including the amino
 R: Tanaka, Y.; Ashikari, T.; Nakamura, N.; Kiuchi, N.; Shibano, Y.; Amachi, T.; Yoshizumi
 Agric. Biol. Chem. 50, 965-969, 1986
 A>Title: Comparison of amino acid sequences of three glucoamylases and their structure-f
 A:Reference number: A90023
 A:Contents: annotation; homology; predicted secondary structure
 C:Comment: Rhizopus glucoamylase exists in multiple forms, Gluc 1, Gluc 2, and Gluc 3, a
 and degrade raw starch.
 C:Genetics:
 A:Introns: 51/3; 110/3; 129/1; 436/3
 C:Superfamily: Rhizopus glucan 1,4-alpha-glucosidase; glucan 1,4-alpha-glucosidase homol
 C:Keyword: glycosidase; hydrolase; polysaccharide degradation
 F:1-25/Domain: signal sequence #status Predicted <SIG>
 F:1-56-604/Product: glucan 1,4-alpha-glucosidase (Gluc 1) #status predicted <GU1>
 F:1-138/Domain: glucoamylase starch-binding domain homology <SBD>
 F:1-116-604/Product: (or 110-604) glucan 1,4-alpha-glucosidase (Gluc 3) #status predicted
 F:1-59-604/Product: glucan 1,4-alpha-glucosidase (Gluc 2) #status predicted <GU2>
 F:1-166-592/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match 26.1%; Score 724; DB 1; Length 604;
 Best Local Similarity 36.8%; Pred. No. 5e-41;
 Matches 165; Conservative 71; Mismatches 170; Indels 42; Gaps 10;

QY 25 ATLDSEWLSNEATVARTAILNNIGADGAWSGADSGIIVASPTNDPDIYFYTWTRDSGLV 84
 Db 168 STISSWIKQEGISRAMLRINP-----PGSATGFIASLSTAGCDIYYAMTRDAALTS 222
 QY 85 KTLVLDLFR--NGDTSLSTIENYISAQAIIVOGISNPSGDLSSGAGIGEPKENVDETAYT 141
 Db 223 NVLYEYNTTLTSGNKTLINLVLDKYVFPVSKQSTSTVCN-----CLGEKFPNDGSGYT 276
 QY 142 GSWGRQQRGPRALRATNIGFG--QWLLDGYSTARDIYWPLYVRNDLSVVAQVMNQTGYD 200
 Db 277 GAWGRQNDGPARRATTFILFDSYLTQTKDASVYGTLPALFAPKLDVYVNWVNSGCCFD 336
 QY 201 LMEEVNSSFPTIAYOHRALVEGSAFATVAGSS--CSWQDAPEIILCYLQSEFMTSFL 258
 Db 337 LMEEVNCHFTLIMWRKELLGADPAKQNDSTRASTYSTASTIANKISSFWSSNNW 356
 QY 259 ANFDS-----SRSGKADANTL--LGSIHNTDPEAACDSTFQPCSPRALANHEVVD 308
 Db 397 IQVSGSVTGVSKKGLDVSTLLAANLGSV-----DDGFPTPSBKILATAVAVEDS 447
 QY 309 FRSITTLNDGLSDSEAVNAGRPEDTY-----YNGNPWFICTLAALPOLYDALYOMDKQ 363
 Db 448 FASLYPINKNLPSTYLGNSIGRYPEDTYNNGNSQGSWFLAVTGVALYRAIKEMING 507
 QY 364 SLEVTDVSLDFPKALYSDATG--TVSSSSSTYSIYDAVTPADGFVSIYETHAASNGSM 422
 Db 508 GYVSSISLPFFKFPSSATSGKTYVGTSDFNFLAQNIALADRFLSTVQLHAHNGSL 567
 QY 423 SEQYDKSDGEOLSARDLTWSYALTLTAN 450
 Db 568 ABEFRTTGLSGARDLTWSHSLITAS 595

RESULT 13

A54549
 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomycopsis fibuligera)
 N:Alternate names: 1,4-alpha-D-glucan glucosylhydrolase; glucoamylase
 C:Species: Saccharomycopsis fibuligera
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 02-Jun-2003
 C/Accession: A54549; S14596
 R:Hostinova, E.; Balanova, J.; Gasperik, J.
 FEMS Microbiol. Lett. 67, 103-108, 1991
 A>Title: The nucleotide sequence of the glucoamylase gene GLA1 from Saccharomycopsis fib
 A:Reference number: A54549; MID:92137640; PMID:1840532
 A:Accession: A54549
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-519 <HOS>
 A:Experimental source: KZ
 A>Note: sequence inconsistent with nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBI:78827, NCBI:78828)
 R:Hostinova, E.; Gasperik, J.
 submitted to the EMBL Data Library, March 1991
 A>Description: Nucleotide and deduced amino acid sequence of the glucoamylase gene from
 A:Reference number: S14596
 A:Accession: S14596
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-271, 'T', 273-519 <HO2>
 A:Cross-references: EMBL:X58117
 C:Genetics:
 A:Gene: GLA1
 C:Superfamily: glucan 1,4-alpha-glucosidase; yeast type; glucan 1,4-alpha-glucosidase hom
 C:Keyword: glycosidase; hydrolase; polysaccharide degradation
 F:1-506/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match 24.4%; Score 676.5; DB 1; Length 519;
 Best Local Similarity 36.1%; Pred. No. 6.4e-38;
 Matches 175; Conservative 71; Mismatches 176; Indels 63; Gaps 16;

Db 42 RTLETFLEDKOXDVSLYYLLQNTAYPEGQNDGYPGTVIASPSTSNPDYYQWTRDSAIT 101
QY 83 ---VLKTLVDLFRNGDSTLSLTENYISAQAIVOGISNPSG--DLSSGAGLGEKFNVD 137
Db 102 FLTVLSLEED--NNFNTTLAKAVEYINTSYNLTORTSNPGSPDDEHKHGLGEKFNVD 159
QY 138 TAYTGSNGRQORDCPALRATAM-----IGFGMLL---DNGYTSRTATDIWPLVR 184
Db 160 SATYGAAGRPQNDGPPALRAVAISRYLNDVNSLNGKLVLTDSGDINPST-EDYKNIIR 218
QY 185 NDISSVYAQYNNQGTGYDLMEEVNGSSFFTIAYQHRALVEGSAFATAV-----GSSGSCWD 238
Db 219 PDLEFYVGYDSTIGFDLMEENQGRHFTSLVQKAL---AYAVDIKSPDDGDFASYLD 274
QY 239 SQAPEILCYLQSFMTG-----SFIILANFD---SSRSKGDANTLLGSHTFDEAACDD 288
Db 275 STASTLESISGSDGFVNTDVNHTVENPDLIQNSQGLDSATYIGPLTLHD-IGSSSS 333
QY 289 STPOPCSPRALANKEVVDSPRSITYTLNDGLSDSEAVAVGRYPEDTY-----YNGNWF 343
Db 334 TPFVDVDEYVYLOSYYLLLEDNKORYSVNSAY--SAGAAIGRYPEDVYNGDSSSGNFWFL 391
QY 344 CTLLAAEQLYDALYQWMDKQSGLEVY--DVSLDFPKALYSDAT-----GTYSSS 391
Db 392 ATAYAAQVPPKLYVD-AKSNANDITTKINDFFNKITYVDLSTNSGYQSDSVTITSGS 450
QY 392 STYSSIVDAVKTFADEGVSIETHAASNGSMSEQYDKSDGEQLSARDLTWSYALLTANN 451
Db 451 DEFNTVADNLVTFPDSFLQVILDHINDGSLNEDLNRYGTSTAYSLTWSGALLERIR 510
QY 452 RKNV 456
Db 511 LRNRY 515

RESULT 14

S48474 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) SGAL - yeast (Saccharomyces cerevisiae)

N:Alternate names: glucoamylase; protein YIL099w

C:Species: Saccharomyces cerevisiae

C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 02-Jun-2003

C:Accession: S48474; C26877; E27284

R:Bowman, S.; Churcher, C.
Submitted to the EMBL Data Library, September 1994

A:Reference number: S48474

A:Accession: S48474

A:Molecule type: DNA

A:Residues: 1-549 <BOM>

A:Cross-references: GB:Z47047; EMBL:Z38125; NID:g603997; PID:g763247; GSPDB:GN00009; MIP

R:Yamaashita, I.; Nakamura, M.; Fukui, S.

J. Bacteriol. 169, 2142-2149, 1987

A:Title: Gene fusion is a possible mechanism underlying the evolution of SNA1.

A:Reference number: A91831; MUID:87194600; PMID:3106330

A:Accession: C26877

A:Molecule type: DNA

A:Residues: 1-503,507,'W',513-514,'TG',516 <YAM>

A:Cross-references: EMBL:M1616; NID:g172592; PIDN:AAA5042.1; PID:g172593

R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.

FEBS Lett. 239, 179-184, 1988

A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar

F:77-536/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match 21.5%; Score 597; DB 1; Length 549;

Best Local Similarity 30.4%; Pred. No. 1,6e-32;

Matches 153; Conservative 89; Mismatches 202; Indels 60; Gaps 14;

QY 1 MFRSLALSGLVCTGLANVSKRATLDSWLSNEATYARAILNNTGADGAWSGADSGI 60
Db 56 VOARDVAVNMGTVYVD-SNGAMDSALEEWLQCKKYSIEKTEENIGPSAVYPS-ISP 113
QY 61 VVSPSTNDPDYFTWTRDGLVTLKTLVDLFRNGDSTLSLTENYISAQAIVOGISNPSG 120
Db 114 VVSPQTHDYTYQMTIRDSALITNSIVS--HSAGPALETLQYLVNSPHLRNN--- 167
QY 121 DLSSGAG-----LGEKFNVDETAYTGSWGRPORDCPALRATAMIGFGQMLDNG--- 170
Db 168 TLGAGIGYTDVVALGDPKNNVNDTAFTEDEWGRPQNDGPPALRAISAILIKITDYK 227
QY 171 -----YTSTATDIWPLVRNDSLVAQYNNQGTGYDLMEEVNGSSFFTIAYQHRALVEGSA 225
Db 228 GARYPQSTP-DLFEDIVRDLRFILDHMNSGFDLMEEVNGMHPFTLLVQLSAVDKSL 286
QY 226 FATAVGSSGSCWD---SQAPEILCYL---QSFMTGSFIILANF-----DSSRSK 270
Db 287 YFNASERSSPFVEELNQRDIDSKFLVDPANGFINCKY--NYIVGPMIADTLRSGLDI 343
QY 271 NTLGSIHTFDEPACDDSTFQPCSPRALANKEVVDSPRSITYTLNDGLSDSEAVAVGRY 330
Db 344 STLLAANTVADAPSA-SHLPEFDINDPAVLNTLHMLHMSIYIPINDSSKNAAGIALGRY 402
QY 331 PEDTY-----YNGNPFWCTLLAAEQLYDALYQWMDKQSGLEVYDVSDFPKALYSDATG 385
Db 403 PEDYDGYGFGEGNPMWLATCTASTLYLYHISEQHDLVVPMNDCSNAPWSELVFS 462
QY 386 TYSS-----SSSTYSSIVDAVKTFADEGVSIETHAASNGSMSEQYDKSDGEQ 433
Db 463 NLITLGNDEGYLLIFERTPAFNQTIQIRPLADSLFKKAHGTGDELSQGNKYTFM 522
QY 434 LSARDLTWSYALLTANNRNSV 457
Db 523 QGAQHLTWSYTSFWDAYQIRQEV 546

RESULT 15

ALBYG glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - yeast (Saccharomyces cerevisiae)

N:Alternate names: 1,4-alpha-D-glucan glucohydrolase; glucoamylase

C:Species: Saccharomyces cerevisiae

C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 05-Sep-1997

C:Accession: A21896; A23470

R:Yamaashita, I.; Suzuki, K.; Fukui, S.

J. Bacteriol. 161, 567-573, 1985

A:Title: Nucleotide sequence of the extracellular glucoamylase gene STA1 in the yeast Sac

A:Reference number: A21896; MUID:85104778; PMID:3918017

A>Note: S. diastaticus

A:Accession: A21896

A:Molecule type: DNA

A:Residues: 1-778 <YAL>

A:Experimental source: strain 5106-9A; ATCC 60709

R:Yamaashita, I.; Suzuki, K.; Sakuro, F.

Agric. Biol. Chem. 50, 475-482, 1986

A:Title: Proteolytic processing of glucoamylase in the yeast Saccharomyces cerevisiae.

A:Reference number: A23470

A:Accession: A23470

A:Molecule type: protein

A:Residues: 1-65 <YAZ>

C:Genetics:

A:Gene: STA1

C:Superfamily: yeast glucan 1,4-alpha-glucosidase STA1; glucan 1,4-alpha-glucosidase hom

C:Keywords: extracellular protein; glycoprotein; glycosidase; hydrolase; polysaccharide

F:33-778/Product: glucan 1,4-alpha-glucosidase #status predicted <MPT>

F:345-778/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
 F:46,319,333,425,434,445,524,557,656,661,731,752/Binding site: carbohydrate (Asn) (coval
 F:635/Active site: Asp #status predicted

Query Match 18.8%; Score 522; DB 1; Length 778;
 Best Local Similarity 30.5%; Pred. No. 2,9e-27;
 Matches 139; Conservative 80; Mismatches 177; Indels 60; Gaps 15;

QY 1 MSFRLSLALSGVCTGLANVLSKRALDLSWLSNEATVATRAILNNIGADGAWSGADSGI 60
 DB 324 VQLRDVLMKNGVTVVD-SNGAMDSPLLEWLRQKKVSIIRIFENIGPSAVYFS-ILPGV 381
 QY 61 VVASPEIDNDPEFYTWTRDSGLVLTLDLFRNGDTSLSTIENYISAQAVOGISNPG 120
 DB 382 VIASPGQTHPDYFYQWIRDSALTNISIVS--HSADPA-ITTLQYLNVSPHLQRTNN--- 435
 QY 121 DLSSGAG-----LGEPRNVDEIATYTGSGWRPQRDGPALRATAMIGFQWMLDNG--- 170
 DB 436 TLGAGIGYTDVVALDGPKNVNDNTAFTEPWGRPQNDGPALRSIALIKIIDYIKOSGTDL 495
 QY 171 -----YTSTATDIWPLVRNDLSVAQYNNQGTGYDLMEEVNGSSFFTIAYQHRALVBSGA 225
 DB 496 GAKYFPOSTA-DIFDDIVRWDLRFIIDHMNSGFDLMEEVNGMFFTLVQLSAVDRLS 554
 QY 226 FATAVSGSCWCD---SQAPEILCYL---QSFWTGSFILANF-----DSSRSQKDA 270
 DB 555 YFNASRRSPFVBELAQTRDISKFLVDPRANGFINCKY---NYIVETPMIADTLRSLDI 611
 QY 271 NTLLSGIHFTDPACDDSTFQPCSPRALANKEVDSFRSIYTLNDGSLDSEAVAVGRY 330
 DB 612 STLAAVTYHADPSA-SHLFPDINDPAVLNTLHLMHMSIYIPINDSSKNATGIALGRY 670
 QY 331 PEDTY-----YNGNPWFCLTAAEQLYDALYQMDKGSLEYVDVSLDFPKALYSDATG 385
 DB 671 PEDVYGYGVGEENPWLATCAASTLYLYRHSIEQHDLVVPNNDCSNAPFSELVFS 730
 QY 386 TVSS-----SSSTYSSIVDAVKTFADGFV 409
 DB 731 NLTLGNDEGYLLENTPAFNQTIQIRPLADSL 766

RESULT 16

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) STRA2 - yeast (Saccharomyces diastaticus)
 N/Altenate names: glucanamyase II
 C/Species: Saccharomyces diastaticus
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 R/Accession: J00474; S35895
 R/Lambda: M.G.; Precorius, I.S.; Sollitt, P.; Marmur, J.
 Gene 100, 95-103, 1991
 A/Title: Primary structure and regulation of a glucanase-encoding gene (STRA2) in Sac
 A/Reference number: J00474; M01D:9127626; PMID:2055484
 A/Molecule type: DNA
 A/Accession: J00474
 A/Residues: 1-767 <LAM>
 A/Cross-references: GB:M0650; NID:9172733; PIDN:AAA3107.1; PID:9172734
 A/Note: the authors translated the codon CCG for residue 337 as Ala, CCG for residue 364
 R/Kit, K.; Beijer, G.; Lee, S.Y.; Knudsen, F.; Mattoon, J.R.
 A/Description: Cloning of a new allelic variant of a Saccharomyces diastaticus glucoamyl
 A/Reference number: S35895
 A/Molecule type: DNA
 A/Accession: S35895
 A/Residues: 1-163, 'A', 164-622, 'D', 624-767 <KIM>
 A/Cross-references: EMBL:M0490; NID:9172735; PIDN:AAA20560.1; PID:9172736
 C/Genetics:
 A/Genes: STRA2; DEX1
 C/Superfamily: Yeast glucan 1,4-alpha-glucosidase STRA1; glucan 1,4-alpha-glucosidase hom
 C/Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:334-767/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
 Query Match 18.8%; Score 521; DB 1; Length 767;
 Best Local Similarity 30.5%; Pred. No. 3.3e-27;

Matches 139; Conservative 80; Mismatches 177; Indels 60; Gaps 15;

QY 1 MSFRLSLALSGVCTGLANVLSKRALDLSWLSNEATVATRAILNNIGADGAWSGADSGI 60
 DB 313 VQLRDVLMKNGVTVVD-SNGAMDSPLLEWLRQKKVSIIRIFENIGPSAVYFS-ILPGV 370
 QY 61 VVASPEIDNDPEFYTWTRDSGLVLTLDLFRNGDTSLSTIENYISAQAVOGISNPG 120
 DB 371 VIASPGQTHPDYFYQWIRDSALTNISIVS--HSADPA-ITTLQYLNVSPHLQRTNN--- 424
 QY 121 DLSSGAG-----LGEPRNVDEIATYTGSGWRPQRDGPALRATAMIGFQWMLDNG--- 170
 DB 425 TLGAGIGYTDVVALDGPKNVNDNTAFTEPWGRPQNDGPALRSIALIKIIDYIKOSGTDL 484
 QY 171 -----YTSTATDIWPLVRNDLSVAQYNNQGTGYDLMEEVNGSSFFTIAYQHRALVBSGA 225
 DB 485 GAKYFPOSTA-DIFDDIVRWDLRFIIDHMNSGFDLMEEVNGMFFTLVQLSAVDRLS 543
 QY 226 FATAVSGSCWCD---SQAPEILCYL---QSFWTGSFILANF-----DSSRSQKDA 270
 DB 544 YFNASRRSPFVBELAQTRDISKFLVDPRANGFINCKY---NYIVETPMIADTLRSLDI 600
 QY 271 NTLLSGIHFTDPACDDSTFQPCSPRALANKEVDSFRSIYTLNDGSLDSEAVAVGRY 330
 DB 601 STLAAVTYHADPSA-SHLFPDINDPAVLNTLHLMHMSIYIPINDSSKNATGIALGRY 659
 QY 331 PEDTY-----YNGNPWFCLTAAEQLYDALYQMDKGSLEYVDVSLDFPKALYSDATG 385
 DB 660 PEDVYGYGVGEENPWLATCAASTLYLYRHSIEQHDLVVPNNDCSNAPFSELVFS 719
 QY 386 TVSS-----SSSTYSSIVDAVKTFADGFV 409
 DB 720 NLTLGNDEGYLLENTPAFNQTIQIRPLADSL 755

RESULT 17

glucanase (EC 3.2.1.-) - Methanococcus jannaschii
 C/Species: Methanococcus jannaschii
 C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 07-Mar-2003
 R/Accession: A64501
 R/Built, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirsch, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
 A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A/Reference number: A64300; M01D:96337999; PMID:8688087
 A/Accession: A64501
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-615 <BL>
 A/Cross-references: GB:U67601; GB:L77117; NID:92826439; PIDN:AAB99630.1; PID:91592211; T
 C/Genetics:
 A/Map position: FOR1584527-1586374
 A/Start codon: GTG
 C/Superfamily: Sulfolobus type glucan 1,4 alpha glucosidase
 C/Keywords: glycosidase; hydrolase

Query Match 8.0%; Score 221.5; DB 2; Length 615;
 Best Local Similarity 23.1%; Pred. No. 3.6e-07;
 Matches 101; Conservative 46; Mismatches 125; Indels 165; Gaps 20;

QY 59 GIVASPEIDNDPEFYTWTRDSGLVLTLDLFRNGDTSLSTIENYISAQAVOGISNP 118
 DB 284 GGIAPSL-HPDYRYWGRD-GSYISIALDLF-----GIRNIPDRFFERMSKIQ----- 331
 QY 119 SGLSSGAGIGEPKRVDEIATYGSW-----GRPQRGSPALRATAM-----IGFQWML 167
 DB 332 -----NAD-----GSLQNYVYNGK-----RLTAIQTDQIGSLTMM 364
 QY 168 DNGYTSTATDIWPLVRNDLSVAQYNNQGTG-----YDLMEEVNGSSFF 211

Db 934 ADALGDGAENVADGTTTAPPTTIANEYNNVGDALDALLMDDETANGAGAYNASH 993
Qy 406 DGFVSIIVETHAASNGSMSE-QYDKSDGEOLSA-----RDLTWSYAAL 446
Db 994 DGAASITIN--VANGSISESTDAVNGSQATNMMEQONTQIINOAGTDTITVIOENG 1051
Qy 447 LTAANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTWSPSIIVATGCTTTTATPT 506
Db 1052 AGINVTYRTNDGGLAFNDASAGVGAT-----AIG-VNSVAKGDSVAIGGGSVDVDTGI 1105
Qy 507 GSGSVTSTKTTTATKSTSTTRSGM 532
Db 1106 ALGSSSVSRVIAKGSRDTSITENG 1131

RESULT 20

G90250
glucan 1,4 alpha glucosidase (Glucosylase) [imported] - Sulfolobus solfataricus
C/Species: Sulfolobus solfataricus
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 07-Mar-2003
C/Accession: G90250
R/she, Q. Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.U.; Chan-
Jong, I.; Ueffing, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to Genbank, April 2001
A/Description: Sulfolobus solfataricus complete genome.
A/Reference number: A99139
A/Accession: G90250
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-622 <KUR>
A/Cross-references: GB:AE006641; NID:g13814174; PIDN:AAK41262.1; GSPDB:GN00155
C/Genetics:
A/Gene: S800990
C/Superfamily: Sulfolobus type glucan 1,4 alpha glucosidase

Query Match 6.2%; Score 172; DB 2; Length 622;
Best Local Similarity 21.7%; Pred. No. 0.00078;
Matches 97; Conservative 56; Mismatches 165; Indels 130; Gaps 19;

Qy 37 VARTALINIGADGAMVSGADSGIIVASPTNDPDYFTYTRDSGLVLTGLVDFRNGDT 96
Db 268 VSLFVIRNHDVNGSIASDSFV-----KTYGDSYQYCPRDAIAYAL-DLAGYKEL 322
Qy 97 SL--LTIENYIAQAIVQGISNPSGDLSCAGLGEKFNVDATYTGSGRP-QNDGPA 153
Db 323 ALHGFQFISIANSEGFVHYKYNPNTLAS--MHPWY-----YKGRYIPLOEDETA 373
Qy 154 LRATAMIGFQMLDNGYTSTATDI-----VPLVRNDLSYVAQYMNQGTGYDWE 203
Db 374 LEV-----WAIASHY-EKYEDIDELPLKYKFKPALKFMMSFMEGLPKRSPDLME 424
Qy 204 EVNGSFFTIYAVHRLAVEGSAFATVAGSSCWCDSQAPILCYLOSFWTGSFILANFDS 263
Db 425 ERYGHIHYTVSYGVALTKGAKLAYDGD-----ILSEDL 461
Qy 264 SRGKANTILGS-----IHTPDEP-----AACDSTFQPCSPRAL-ANHKEVDSFRS 311
Db 462 DTSLGLKGMVLKMTYNGRFVRRIIDENNODLTVDSSLYAPFFEGFLVANDKIMINTINE 521
Qy 312 I---YTLNDGLSDSEAVAVGRYPEDTY-----NGNPFELCTLLAAEQLYALQWMDKGS 364
Db 522 IESRLVINGGIT-----RIENDMYQRKKQPNPIITTL-----W----- 556
Qy 365 LEVTVDSLDFKALYSDAATGYSSSSSTYSIIVDAVKTADGVSIVETHAASNGSMSE 424
Db 557 -----LSEYATIND--KNANEXYIKVIRALPTGFLPE 589

Qy 425 QYDKSDGEOLSAADLTYSYALLTANR 452
Db 590 QVDPETFEPTSVPLVWSHAFFITAINK 617

RESULT 21
T33369
hypothetical protein H02P09.3 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C/Accession: T33369
R/Giesel, C.; Harmon, G.
submitted to the EMBL Data Library, July 1998
A/Description: The sequence of C. elegans cosmid H02P09.
A/Reference number: Z21330
A/Accession: T33369
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1275 <GEI>
A/Cross-references: EMBL:AF07538; PIDN:AA064622.1; GSPDB:GN00028; CESP:H02P09.3
A/Experimental source: strain Bristol N2; clone H02P09

Query Match 6.2%; Score 170.5; DB 2; Length 1275;
Best Local Similarity 22.4%; Pred. No. 0.0027;
Matches 129; Conservative 87; Mismatches 276; Indels 85; Gaps 19;

Qy 1 MSFRSLALSGVCTGLANVISKRALDLSWLSNEATVA-RTALINIGADGAMVSGADSG 59
Db 673 VSGSVSTVSGSTESVSGASTVSGSTGTSVSDSTISDTNAPGSTESTVTSVS 732
Qy 60 IVVASPTNDPDYFTYTRDS--GLVLTGLV-LFRNGDTSLTIENYIAQAIVQGI 115
Db 733 TVSGSTGTPSTMSASTGNTPGSTESTITDGSIVSGSGTSGTNNCGSIDSTTGI 792
Qy 116 SNDSG-DLS--SGAGLGEKFNVDATYTGSGWRPQRDGPALATAMIGFQMLDNGYT 172
Db 793 STVSGSLSTISGTSVSGSSDMVTGTSSTPSGSTESTVSGASTMS-----P 842
Qy 173 STATDVLWPLVRNDLSYVAQYMNQ--TYDLMEEVNGSFFTIYAVQH-----R 218
Db 843 STGSSVETSGSVSTVSGSTSSSTTQ--STVESSVSTVSSSTISQSTTTGE 899
Qy 219 ALVEGSAFATVAGSSCWCDSQAPILCYLOSFWTGSFILANPDSRGDANTLGSIH 278
Db 900 STVFGSTGSTRATSSMSASTGTDTPGSTESTITGTVG--ESTVSGSTGTTTIG-- 955
Qy 279 TFDPEACDSTFQF--CSPRALANHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTY 335
Db 956 -----STISESTWTVGVSTGSTITGESVSGSTRATVGTESVSGSTESTVSGSTEST- 1009
Qy 336 YNGNPFELCTLLAAEQLYALQWMDKGSLEVTVDSLDFKALYSDAATGYSSSSSTYS 395
Db 1010 ----FTVPSTVSGS-----TGSTVTGTESTVSGSTASTSSGSTSSSTEAGSTVS 1053
Qy 396 SIYDAVKTADGVS-----SIVETHAASNGS--MSRQYDKSDGEOLSAADLTYSVA 444
Db 1054 GGSASTVTSSTGSGTSGESTVSGSTVSGSTGSTITGTSTVSGSTESTVTSBVS 1113
Qy 445 ALITANNRR-----NSVPASWGETSASVPGTCAATSAIGTYSVT--VTSWPSIVAT 496
Db 1114 SVSTVSGNNGSTITGTSVSGSTGSGESTILESSVSTVSGSTITDGSSTRSSVST 1173
Qy 497 GGTITATTPG--SGSVTSTKTTATKSTSTTRSG 531
Db 1174 VSASTESTVSGSSASIGSTNTPDSTESTISGTTISG 1210

RESULT 22

B60754
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) S - creosote fungus (ATCC 20495) (fragments)
C/Species: Amorphotheca resinae, Hormoonia resinae (creosote fungus)
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 02-Jun-2003
C/Accession: B60754


```
Qy 262 DSSRSGKANTLLGSIHTFDPPEACDSTFQPCSPALANHKEVDSFRSIYTLNDGLSD 321
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 506 ATSSSSSDATTTSSSGCT--TSSSDATSSSDATTTSSSGCTATSSSDATTTSSSGTAT 563
Qy 322 SEVAVAGRPEEDTYVNGNPFLLCTLAAEOLYALQWMDQG---SLEVTVDLDPFKA 377
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 564 SSSDVTTSSSEGT-----ATSSSDATTTSSSGCTTSSSDATTTSSSGCTAT 609
Qy 378 LVSDATGTYSSTSSSVSDVADKTFADGFSIVETHAASNGSMSEQD---KSDGRQ 433
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 610 TSSDATTSS--SSSEGTSSSDATTTSSSD--VTTSSSEGTATSSSDATTTSSSEGT 663
Qy 434 LSARDLTWSYAAALLTANRRNSVVPASWG-----ETSAASYGTCAATSAIGTYSVT 486
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 664 TTSSDATTTSSSGCTTSSSDATTTSSSGCTATSSSDATTTSSSGCTATTTSSSG 723
Qy 487 VMSWPSIVATGCTTTATPTGSGSVTSTKTTATASKTSTTRS 530
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 724 TATSSDATTTSSSGCTTSSSDATTTSSSGCTATTTSSSDATTTSSSG 767

RESULT 25
A25547
ice nucleation protein - Pseudomonas fluorescens
C:Species: Pseudomonas fluorescens
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 26-Aug-1999
C:Accession: A25547
R:Warren, G.; Corotico, L.; Wolber, P.
Nucleic Acids Res. 14, 8047-8060, 1986
A:Title: Conserved repeats in diverged ice nucleation structural genes from two species
A:Reference number: A25547; MUID:87040772; PMID:3774551
A:Accession: A25547
A:Molecule type: DNA
A:Residues: 1-1210 <MAX>
A:Cross-references: GB:X04501; NID:g45507; PIDN:CAA28186.1; PID:g45508
C:Genetics:
A:Gene: hmw
C:Superfamily: ice nucleation protein

Query Match 5.7%; Score 159; DB 2; Length 1210;
Best Local Similarity 22.2%; Pred. No. 0.015;
Matches 128; Conservative 72; Mismatches 246; Indels 130; Gaps 25;

Qy 8 ALSGLVCTGLANVSRATLDSWLSNEATVARTAILNMGADGAWMSGADSGIVASPT 67
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 165 AAYGSLTGT---ANOSQLAGYGTETAGDSSTLTAAGYSTG--TSGDSSTLTAAGYST 218
Qy 68 DNPDPFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAVOGISNPSGDLSSG-- 125
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 219 G-----TAGSDSLTAAGYSTGTAGDSSLTA---CYGSTQTAQVGSNLTAGYGTGTA 269
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 126 -----AGLGEKENVDETAATYATGWSGRPQRDGPALRATAMIGFGWLLDNGYTATDI 178
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 270 GPDSSLTAAGYSTGTAGGESSLTAAGYSTG---TAQVGS---LTAAGYSTGT-- 316
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 179 WNPVANDLSYVQVYNNQGYDLMEEVNGSFTTIAVQHRALVEGSAFAFAVSSGSKWD 238
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 317 ----AGSDSSLTA-----GYGSTQTAAGDSSLTAAGYSTQTAQVGSNLTAGYGTGTA 366
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 239 SQAPELLCTYQSFMTG--SFILANPDSRS--GKDANTLLGSIHTFDPPEACDSTFQ 293
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 367 PDBSLTAAGYSTGTAGGESSLTAAGYSTGTQTAQVGSDLTAAGYSTGT---AGSDSSLTA 422
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 294 C--SPRALANHKEVDSFRSIYTL-----NDGLSDEAVAVGRYPEDTYVNGNPN 341
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 423 YGSTQTAAGESSLTAAGYSTGTQTAQVGSDLTAAGYSTGTAGSDSLTAAGYSTGTAGD-- 480
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 342 FLCTTAAAEOLYALQWMDQGLEVTVDLDPFKALYSDAAGTSSSSSTYSIVDAV 401
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 481 --SSLTAG-----YGSTQTAQVGSDLTAAGYSTGTAGSDSLTAAGYSTG----- 521
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 402 KTFADGFSIVETHAASNGSMSEQDKSDGQLSARDLTWSYAAALLTANRRNSVVPASW 461
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
Db 522 -----TGT-AGDSSLTAAGYSTGTQTAQVGS--DLTAAGYSTGTAGS--DSSLTAAG 567
Qy 462 GENSASSVPGTCAATSAIGT-----YSSV--TVTSWPSIVATGCTTTATPTG 507
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 568 GGTQTAG--GDSSSLTAAGYSTGTQTAQVGSNLTAGYGTGTAAGSDSLTAAGYSTQTA--GG 623
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 508 GGSVSTSTSKTTATVASK-----TSTTTRSGMSL 534
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 624 DSSLTAAGYSTGTAGGSLTLAGYGSTQTAQVGS 659

RESULT 26
F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90073
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.; Shibata, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90073
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2271 <KID>
A:Cross-references: GB:BA000018; PID:g13702612; PIDN:BA843752.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2447

Query Match 5.7%; Score 159; DB 2; Length 2271;
Best Local Similarity 20.4%; Pred. No. 0.036;
Matches 116; Conservative 105; Mismatches 263; Indels 84; Gaps 17;

Qy 21 ISKRATLDSWLSNEATVARTAILNMGADGAWMSGADSGIVASPTDNPDPFYTWTRDS 80
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 748 VTRNSMDS--VTSSTGTOQSGSVSTSKADSQASSTSSGSIYVSTASTSKTSVLSBS 806
Qy 81 GLVLKTLVDLFRNGDTSLSSTIENYISAQAVOGISNPSGDLSSGAGLGEKENVDETA 140
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 807 VSASKSLSTSESNVS--STSTLVNSQSV---SSSMGGSVSKSTSLSDSISNSTEK 861
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 141 TSGWGRPQRDGPALRATAMIGFGWLLDNGYTATDIWPL-----VANDLSYVQVYNN 195
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 862 SSSLSTSD--SLRSTSLSDLSMSTSGSKSLSTSSGSSSTASLSLSDSTSNAI 919
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 196 QGYDLMEEVNGSFTTIAVQHRALVEGSAFAFAVSSGSKWDSPQAPFI-----LC 246
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 920 STSTLSSEASTSDSISIS-----NSINSGASSTSKSDSOSTSISLSTSDSKMS 970
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 247 YLQSF-----WTGSPFLA-----NPDSSRGKANTLLGSIHTFDPPEACDSTFQ 292
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 971 TSESLDSTSTSGSVGSLIAASQSVSTSDSMSTSIIVSDSISTSGSLASDCKMS 1030
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 293 PCSPRALANHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYVNGNPFLLCTLAAEOL 352
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1031 VSSSMSTSGSSTSESLSDSOSTSD--SDSKSLT-----STSGSGSTSTSTSTASVPT 1083
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 353 YDALYQWMDQGLEVTVDLDPFKALYSDAAGTSSSSSTYSIVDAKTFADGFSIV 412
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1084 SES---QSTSGSVASQSDSMGSLSTSFSDTSDSKASASTASSESLSQASSTSGSVST 1140
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 413 ETHAASNG-----SMSEQDKSD--GEQLSARDLTWSYAAALLTANN 451
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1141 TSLSTNSERTSTGVS DSTSLSTSESDSISESTSDSISLASEST--SISLSEMS 1198
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 452 RNSVVPASWGERTASAVGTCAATSAIGTYSVVTSPSPVATGCTTTA---TPTGS 508
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1199 TSDSESQASAPFISELSESTSESTSESVSSSTSESTSLSDSTSESGSTSLSNSTSGS 1258
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 509 GSV---TSTSKTTATASKTSTTRGMS 533
  ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Db 1259 ASISTSTISSESTFKSESVSTSLSMS 1286

RESULT 27

T39174 hypothetical Serine/threonine repeat containing protein [imported] - fission yeast (Schl

C/Species: Schlizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000

C/Accession: T50375; T39172; T39173; T39174; T39366

R/Conor, R.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, February 1998

A/Reference number: Z21832

A/Accession: T50375

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1283 <CON>

A/Cross-references: EMBL:AL021815; PIDN:CAB61533.1; GSPDB:GN00067; SPDB:SPBCB4.70C

A/Experimental source: strain 972h-; cosmid c8E4

A/Accession: T39172

A/Molecule type: DNA

A/Residues: 785-1283 <CO2>

A/Cross-references: EMBL:AL021815; PIDN:CA17000.1

A/Note: this is an interim translation for a sequence replaced by GenBank/EMBL

A/Accession: T39173

A/Molecule type: DNA

A/Residues: 'ME', '179', 'PLV', '183', 'W', 'KL', '556-761', 'HRGSS' <CO3>

A/Cross-references: EMBL:AL021815; PIDN:CA17001.1

A/Note: this is an interim translation for a sequence replaced by GenBank/EMBL

A/Accession: T39174

A/Molecule type: DNA

A/Residues: 1-555, 'S' <CO4>

A/Cross-references: EMBL:AL021815; PIDN:CA17002.1

A/Note: this is an interim translation for a sequence replaced by GenBank/EMBL

R/Mod: V.; Rajandream, M.A.; Barrell, B.G.; Oliver, K.; Harris, D.

submitted to the EMBL Data Library, March 1999

A/Reference number: Z21848

A/Accession: T39366

A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: DNA

A/Residues: 1-896, 'G', '898-904', 'I', '906-1283' <MOO>

A/Cross-references: EMBL:AL035675; PIDN:CA38695.1; GSPDB:GN00067; SPDB:SPBC1289.15

A/Experimental source: strain 972h-; cosmid c1289

C/Genetics:

A/Gene: SPDB:SPBC1289.15; SPDB:SPAC8E4.07C

A/Map position: 2

C/Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

Query Match 5.7%; Score 158; DB 2; Length 1283;

Best Local Similarity 21.0%; Pred. No. 0.019;

Matches 118; Conservative 76; Mismatches 244; Indels 124; Gaps 24;

QY 5 SLALSLGVLCTGLANVISKRAITLDSWISNEA-TVARAILNIGADAM-VSGADSGIV- 61

DB 180 SLIPHSITSLSTIVSINDTSASLSKTTSPRAGITTEIVSGSVGTSTFPASGTTSTVE 239

QY 62 VASPTDNPDPFYFTWRDSCGLVLTVDLPFRNGDTSLSTIENYISAOAIVOGISNSGD 121

DB 240 VVERP-----AGITTEIV-----SGSVGTSTFPANGTSGIVEVEPFRAG- 281

QY 122 LSSGAGLGEPRKPNVDEFAIVYSGWGRPQRDGPALRATAMIGFGOMLDNGYTTATDIWMP 181

DB 282 -----TITTEIVSGS-----VGYSFPFANGTTSIGVEVEP 313

QY 182 LVANDLSYVAQYNNQOTGYDLMEFVNGSSFTIANQH-----RALVEGSAFAT---AV 230

DB 314 TACTITTEIVS--GSVGYTSTFPANGTTSIGVEVEPFRAGITTEIVSGSVGTSTFPAS 371

QY 231 GSSCSWCDQAPETILCYLQSFMTGS-FLANFDSRSRGKADNTLLGSIHFDPEAACDS 269

DB 372 GTTSGIVVEVPEPFRAGITTEIVSGSKAFSTFPAN-----GTTSGIVEVEPFRAGITTK 425

QY 290 TFPQCPSPRALANHKEVVDSPRSIYTLNDGLSDSEAV--AVGRYPEDTYNGNFWPLCTL 346

Db 426 TIVSGS-----KTFTSTFPANGTTSIGVEVEPFRAGITTE-TIVSGSVGTSTFP 473

QY 347 AAAEQLYDALYQWDKQGSLEVTVDVSLDPFKALYSDAATGYSSSSSTYSIYAVATFPAD 406

DB 474 PA-----NOTTSGIVEVEP-----AGITTEIVSGSKTFTSTFPASGT-TS 515

QY 407 GFSIVTEHAANSNGNSEQY---DKSDGEQLSARDLTWGYALL--TANNRNSVVPASW 461

DB 516 GTVEVEPFRAGITTEIVSGSKAFSTFPANGTTSIGVEVEPFRAGITTEIVSGSV 572

QY 462 GFTSASVGTCAATTAIGYSSVVT-----SWPSIVATGGTTTATPTGSG 509

DB 573 GYTSFPASGTTSGIVEVEPFRAGITTEIVSGSVGTSTFPASGTTSGIVEVEPFRAG 631

QY 510 SVTSTKTATATASKTSTTTRSG 531

DB 632 TTTETI-VSGSVGTSTFPASG 652

ice nucleation protein - Erwinia herbicola

C/Species: Erwinia herbicola

C/Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 26-Aug-1999

C/Accession: J00188; S35273

R/Warren, G.; Corotico, L.

Gene 85, 239-242, 1989

A/Title: The consensus sequence of ice nucleation proteins from Erwinia herbicola, pseud

A/Reference number: J00188; MUID:90152370; PMID:2515997

A/Accession: J00188

A/Molecule type: DNA

A/Residues: 1-1258 <WAR>

A/Cross-references: EMBL:M26382; NID:9148419; PIDN:AAA24823.1; PID:9148420

A/Experimental source: strain M1

R/Gurian-Sherman, D.; Lindow, S.E.; Panopoulos, N.J.

Mol. Microbiol. 9, 383-391, 1993

A/Title: Isolation and characterization of hydroxylamine-induced mutations in the Erwinie

A/Reference number: S35273; MUID:94018633; PMID:8412688

A/Accession: S35273

A/Status: preliminary

A/Molecule type: protein

A/Residues: 387-536 <GUR>

C/Genetics:

A/Gene: iceE

C/Superfamily: ice nucleation protein

Query Match 5.7%; Score 157; DB 2; Length 1258;

Best Local Similarity 22.1%; Pred. No. 0.021;

Matches 132; Conservative 63; Mismatches 239; Indels 162; Gaps 25;

QY 47 GADGAWVSGADSGIVVSPSTDPDPFYFTWRDSCGLVLTVDLPFRN-----GPTSL 99

DB 299 GADSLIAGYGSQTQAGEESTQTAGYGSTQTAQK-----SLLTAGYSGTQAGDS-- 350

QY 100 STIENYISAQAIYQGISNPSG-----DLSSG-----AGLGEPRKN 134

DB 351 SLIAGYGSQTQAGESSLTAGYSGTQTAQKSDLTLAGYSGTQAGADSLIAGYSGTQTA 410

QY 135 VDETAATGSKWRPQRDGPALRATAMIGFGOMLDNGYTTATDIWMPPLVANDLSYVAQY 194

DB 411 GRESTQTAGYGSQT-----TAQKGSQ-----LTAAGYSGT-----AGDDSLA-- 450

QY 195 NQGYDLMEFVNGSSFTTIAVQHRALVEGSAFATAYGSSCSWCDQAPETILCY-LSQFW 252

DB 451 ---GYSQTQAGESSLTLAGYSGTQTAQKSDLTLAGYSGTQAGSSSLIAGYSGTQTAG 507

QY 253 TGSFIILANFDSRSRGKADNTLL---GSIHFDPEAACDSTFPQCPSPRALANHKEVVD- 308

DB 508 YGSLTLTAGYSGTQTAQNESDLITNGYSGT-----AGANSSLIAGYSGTQTAASVNSVLTAG 563

QY 309 FRSIYTLNDGL-----SDSEAVA-VGRYPEDTYNG-NPWFICTIAAABQ-- 351

Db 564 YGSTQTAAREGSDLTAGYGTGTAGSDSSIIAGYGTQTASVHSSLTAGYGTQTAREQSV 623
QY 352 -----LYDALYQWMDKOSLEVTVDVSLDFPKALYS 380
Db 624 LTTGCGSTGTAGADSSLIAGYGTGTATAGVNSILTAGY-----GSTQTAQEGSLTAGYGS 678
QY 381 DAATGYSSSSSSSYSSIVDAVKTADGVSVIVETHAASNGSMSEQYDKSDG-----431
Db 679 TGTAGADSSLIAGYGS-----TGTAGVNSILTAGYGTGTQTAQEGSLTAGYGTGTAG 732
QY 432 -EQLSN---RDLTWSYALLLTA-----NNRRNSVVPASMGFTSASSVPETCAATSAIGT 481
Db 733 DSSLINGYGTGTATASHSSLTAGYGTGTATAREQSVLTGTGSGTSTAG-----ADSSLIAG 787
QY 482 YSSVTVTWSPIVATG-GTTTAA-----PPTGSGSVTSTKSTTATATASKSTTTRSG 531
Db 788 YGSTGTATGHSILTAGYGTGTATQERSDLTTGSGTSTAGADSSLIAGYGTGTATG 843

RESULT 29

T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C/Accession: T34434
R/Gene: C. Gattung, S.
Submitted to the EMBL Data Library, December 1996
A/Description: The sequence of C. elegans cosmid K06A9.
A/Reference number: Z21525
A/Accession: T34434
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2232 <GEI>
A/Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A/Experimental source: strain Bristol N2; clone K06A9
C/Gene: CESP:K06A9.1a
A/Map position: X
A/Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/

Query Match 5.6%; Score 156.5; DB 2; Length 2232;
Best Local Similarity 23.4%; Pred. No. 0.051;
Matches 113; Conservative 70; Mismatches 228; Indels 71; Gaps 18;
QY 80 SGLVLTLDLPFRNGDLSLTIENTYISQAIVQGISNPSGDISGAGLGEPEKFNDETA 139
Db 1440 SGGTQHTTMSKASSGST---SPSTNSQTGSTVMGSSSTSGVSTSSASTQPMSTSGS 1496
QY 140 YTGSMWRPORDGPALRATAMIGFQWLLDNGYSTATDIWPLVRMDLSVAQYMRQGY 199
Db 1497 SAGSTYASSTASPAASSTAPSSGT-----MSTSGSYVGTI -SSSTTASASQTG- 1548
QY 200 DLMEVNGSFFTIQV-----HRLVGSAPATAVGSSCSW--CDQAPEILCYL 248
Db 1549 ---STYTMGSSSTSGVSTSSASTQPMSTSGSAGSTYASSTAGLVSTVPSSTGM 1605
QY 249 QSFWTSSFLIANDSRSGKADANTLLGSIHTFPDEAACDDSTFQPCSPRALANHKYVDS 308
Db 1606 GSTYSSGT-VGSTISESSTTASASQTGSTVMGSSSTSGVSTSSASTQPMSTSGSSG 1664
QY 309 FRSIYTLNDGLDSEAVAVGRYPEDTYVNGNP---WFLCTLAAEQLYDLYQWMD---Q 362
Db 1665 GSTVASTTGL-----VSTSTVPSSTGTCTMGSTSSGTVGSISSSTRAASASQTGSTM 1719
QY 363 GSLEVTVDVSLDFPKALYSDAATG---TVSSSSSTVSSIYDAVKTADGVSVIVETHAAS 418
Db 1720 GSSSTSGVST-----SSASSGQPMSTSGSSAGSTVVSSTAPAS-----STAPSS 1767
QY 419 NSGSMFOYKSDDEQLSARDLWTSYALLTANNRRNSVVPASG-----ETGASSVPCR 472
Db 1768 TGTMSSTSGTGVSTWSQ-----SSTPAASTTSHGTGTVLGGSSSTSSNQSTSGSSVGS 1822
QY 473 CAATSAIGTYSVTVTWSPIVATGTTTATPTGSGSVTSTKTATA-SKSTTTRSG 531

Db 1823 TVASSTAGLVSTIV---PSSGTGWTGSSGT---VGSSTISESSTTASASQTGSTVWG 1876
QY 532 MS 533
Db 1877 SS 1878

RESULT 30

S07053
ice nucleation protein inaA - Erwinia ananas
C/Species: Erwinia ananas
C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 26-Aug-1999
C/Accession: S07053
R/Abbe, K.; Matabe, S.; Emori, Y.; Watanabe, M.; Arai, S.
FEBS Lett. 258, 297-300, 1989
A/Title: An ice nucleation active gene of Erwinia ananas. Sequence similarity to those of
A/Reference number: S07053; M0ID:90092494; PMID:2599095
A/Accession: S07053
A/Status: not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-1322 <ABE>
A/Cross-references: GB:X17316; NID:G296095; PIDN:CAA35194.1; PID:G296096
C/Superfamily: ice nucleation protein

Query Match 5.6%; Score 155; DB 2; Length 1322;
Best Local Similarity 22.0%; Pred. No. 0.031;
Matches 125; Conservative 69; Mismatches 220; Indels 154; Gaps 27;

QY 47 GADGAWVGADSGIIVASPSDNDPFIYTWTRDSGLVLTVDLPFN-----GDTSL 99
Db 315 GADSSLIAGYGTGTATAGESTGTATAGYGTGTATAGK-----SDLTAGYGTGTAGDSS-- 366
QY 100 STIENTYISQAIVQGISNPSG-----DLSSG-----AGLGEPEKFN 134
Db 367 SLIAGYGTGTATAGESTLTAGYGTGTATAGKGDLTATGSGTSGADSSLIAGYGTGTAT 426
QY 135 VDEATATGSMWRPORDGPALRATAMIGFQWLLDNGYSTATDIWPLVRMDLSVAQYMR 194
Db 427 GRESTTATAGYGTQ-----TAKGSD--LTATAGYGTGT-----AGDSSLIAGY 469
QY 195 N---QTYDVLMEVNGSFFTIQVHRLVGSAPATAVGSSCSWCDQAPEILCYLOS 250
Db 470 STQTAQKGSDEL---TAGYGTSTTAGYESSLIAG-----YGSF-----QT 505
QY 251 FWTGSLIANDSRSGKADANTLL---GSIHTFPDEAACDDSTFQPCSPRALANHKYVD 307
Db 506 AGYGTTLTAGYGTGTATQWESDLITGYGSTST---AGANSSLIAGYGTGTATASVNSVLT 561
QY 308 S-FRSIYTLND-----GLSDSEAVAVGRYPEDTYVNGNPWFLC-----TLAAE 350
Db 562 AGYGTGTATAREGSDLTAGYGTGTATGSDSSIIAGYGTGTATAGDSSLIAGYGTGTAT 621
QY 351 QLYDALYQWMDKOSLEVTVDVSLDFPKALYSDAATGYSSSSSYSSIVDAVKTADGVSV 410
Db 622 SILTAGY-----GSTQTAAREGSDLTAGYGTGTATAGDSSLIAGYGS-----TGTAGVNS 670
QY 411 IV-----ETHAASNGS-MSEQYDKSD-----GEQLSA---RDLTWSYALLLTA 449
Db 671 ILTAGYGTGTATQWESDLTAGYGTGTATAGDSSLIAGYGTGTATASVNSILTAGYGTGTAT 730
QY 450 NNRRNSVVPASMGFTSASSVPETCAATSAIGTYSVTVTWSPIVATG-GTTTAA-----503
Db 731 --QEOSVLTGTTGSGTSTAG-----ADSSLIAGYGTGTATAGVNSILTAGYGTGTATQARESD 783
QY 504 TPTGSGSVTSTKTATATAKSTTTRSG 531
Db 784 LTTGCGSTGTAGADSSLIAGYGTGTATG 811

RESULT 31

SNP50
ice nucleation protein [validated] - Pseudomonas syringae

N,Alternate names: Snomax (TM)
C,Species: Pseudomonas syringae
C,Date: 31-Mar-1988 #sequence_revision 21-Jan-1997 #text_change 15-Sep-2000
C,Accession: A24405
R,Green, R.L.; Warren, G.J.
Nature 317, 645-648, 1985.
A,Title: Physical and functional repetition in a bacterial ice nucleation gene.
A,Reference number: A24405
A,Molecule type: DNA
A,Residues: 1-1200 <GRE>
A,Cross-references: EMBL:X03035; NID:g45828; PIDN:CAA26837.1; PID:g45829
R,Kajava, A.V.; Lindow, S.E.
submitted to the Brookhaven Protein Data Bank, June 1993
A,Reference number: A51242; PDB:1YNA
A,Contents: annotation; theoretical model, residues 490-535
R,Kajava, A.V.; Lindow, S.E.
J. Mol. Biol. 232, 709-717, 1993
A,Title: A model of the three-dimensional structure of ice nucleation proteins.
A,Reference number: A58442; MUID:93360260; PMID:8355267
A,Contents: annotation; theoretical model
C,Comment: Found on the outer membrane of the bacteria, this protein stimulates ice form
C,Superfamily: ice nucleation protein
C,Keywords: tandem repeat
F,208-1151,Region: 8-residue repeats (A-G-Y-G-S-T-L-L-T)

Query Match 5.6%; Score 154.5; DB 1; Length 1200;
Best Local Similarity 22.9%; Pred. No. 0.029;
Matches 144; Conservative 69; Mismatches 254; Indels 163; Gaps 29;

QY 9 LSGIVCTGLANVISKRATLDLSWLSNEATVARTALINNIGADGAWSGADSGIVVASPSTD 68
DB 207 IAGGSGTGTAG-----SDSWL-----VAGYSGTGTAGDSDALTAGYSGTGTARBSNL 254
QY 69 NPYFYTW--RDSGLVLTGLVDFRNGDPSLSTIENYISAQIVOGISNPSGDLSSG- 125
DB 255 TAGGSGTGTAGSDSLIAGYSGTGTSGDSSLTA--GYSGTGTAGSGNLTAGYSGTGT 311
QY 126 -----AGLGEKPFNVDETAAYGSGWRPQ--RDGPALRA-----TAMGFG 163
DB 312 AGSDSLIAGYSGTGTSGDSSLTAAGYSGTGTAEBSNLTAGYSGTGTAGVDSLIAGY 371
QY 164 QWL-----LDNGYSTATDIPVPLVNDLSYVAQYWNQGYD 200
DB 372 STGTSGSDALTAGYSGTGTAEBSNLTAGYSGTGT-----AGDSSLIA-----GYG 419
QY 201 LMEVNSSFFITAV-OHRLVBSAATAVGSS-CWCDSQAPELLCTYQSFWTGSFIL 258
DB 420 STGTSGSDSLTAGYSGTGTAEBSNLTAGYSGTGTAGVDSLIAGYSGTGTSGDALT 479
QY 259 ANPDSRSRGDANTL--LGSIHFDPEACDDSTFQPC-SPRALANHKKEVDSFRSIYT 314
DB 480 AGYSGTGTAEBSNLTAGYSGTGT-----AGADSLIAGYSGTGTSGSESSLTAGYSGTGT 535
QY 315 LNDGL-----SDSEAVAVGRYPEDITYNCPWFLCTLAALQYDALY--Q 358
DB 536 ARGGSTLTAGYSGTGTAGADSSLIA--GYSGTGT-----SGSESSLTAGYSGTQ 582
QY 359 WDQGSLEIVDVLDFPKALYSDAATGTYSSSSSTYSIYDA--VKTFADGPGSIYVTHA 416
DB 583 TAQGGSVLTGSGYGTGTAGAAASNLTTGYSGTGTAGHSSFTIAGYSGTGTAGHKSIL--T 639
QY 417 ASNGSMSEQYDKSD-----GEQLSA--RDLTWSYALTLTA-----NNRRN 454
DB 640 AGYSGTGTADGSLIAGYSGTGTAGSGSSLIAGYSGTGTASTKSMILTAGYSGTGTAKRH 699
QY 455 SVVPASMGETSASVPCTCAATSAIGTYSSVTVTSWPSIVATG-GTTTATP-----TGS 508
DB 700 SDLVYTGSTAG-----SNSSLIAGYSGTGTAGFSLITAGYSGTGTAEKERTSLVAGY 754
QY 509 GSVTSTG-----KTATATASKTSTTT 528

DB 755 GS-TSTAGYSSSLIAGYSGTGTAGESTLT 783

RESULT 32
T41144
hypothetical serine rich protein - fission yeast (Schizosaccharomyces pombe)
C,Species: Schizosaccharomyces pombe
C,Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Dec-2000
C,Accession: T41144; T41591
R,Hilbert, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1998
A,Reference number: Z21973
A,Accession: T41144
A,Status: translated from GB/EMBL/DBJ
A,Molecule type: DNA
A,Residues: 1-1131 <HIL>
A,Cross-references: EMBL:AL031907; NID:g3766362; PIDN:CAA21415.1; PID:g3766363; GSPDB:GN(C
A,Experimental source: strain 972h-; cosmid c18
R,Murphy, L.; Harris, D.; Lyne, R.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A,Reference number: Z21968
A,Accession: T41591
A,Status: translated from GB/EMBL/DBJ
A,Molecule type: DNA
A,Residues: 44-1131 <MUR>
A,Cross-references: EMBL:AL031543; NID:g4239667; PIDN:CAA20837.1; PID:g3581899; GSPDB:GN(C
A,Experimental source: strain 972h-; cosmid c74
C,Comment: For a possible alternative initiator product, see PIR:T42367.
C,Genetic:
A,Gene: SPCC74.07C; SPDB:SPCC18.01C
A,Map position: 3
C,Keywords: alternative initiators

Query Match 5.5%; Score 153; DB 2; Length 1131;
Best Local Similarity 20.4%; Pred. No. 0.034;
Matches 112; Conservative 85; Mismatches 199; Indels 154; Gaps 22;

QY 32 SNEATVARTALINNIGADGAWSGADSGIVVASPSTDNPYFYTWTRDGLVLTGLVDF 91
DB 560 STEALSNLSLSSSTASSTSYSPASSYEVANSBD-----YVS-----QTVSSIT 606
QY 92 RNQDTSLSLSTIENYISAQIVOGISNPSGDLSSGALGEPKRVND-----ET 138
DB 607 ASGTTSTSEI-----VSTPASNSNGSLNGSSFFVNVGSSGQTTPTSSS 654
QY 139 AYVSGSRGPRDGPALRATAMIGFGQWLDNGYSTATDIPVPLVNDLSYVAQYWNQGY 198
DB 655 SITGSGSLKETSSPAL-VSSTVSTSSYVDSSTYNT-----GSSSSDSQSFSGTT 705
QY 199 YD-----LMEEV-----NGSSFFITAVOHRALVEGSAFATAVGS 232
DB 706 YSDPTTTTISEVSSILSPFTMSQSVSRPQSGDAGFTITFISQSSDGETSGYTIS 765
QY 233 SCWCDSQAPELLCTYQSFWTGSFILANPDSRSRGDANTLGSIH-TDPEAACDDSTF 251
DB 766 NSQONSASEPQ-----TAFSSSSSATPTTOSISITSVSSQSNMSSSVS 810
QY 292 QPCSPALANHKKEVDSF-RSIYTLNDGLDSEAVAVGRYPEDITYNCPWFLCTLAAL 350
DB 811 SPILSSNSVSTSTIISIASSTISIPIS-----SIAS 844
QY 351 QLYDA-----LYQMDKQGSLEVTVDLDFPKALYSDAATGTYSSSSSTYSIYDAVKTFA 405
DB 845 SPFDAGFTSIYGTAG-----FSSSF--ALASNSEGASDVLSTTI-----AKPTF- 890
QY 406 DGVSVLVEITHASNGSMSEQYDKSDGEQLSARLUTSYAALLTANRRNSVVPASMGETS 465
DB 891 -----KSTVSGSGTYSIPSS-----SRNEGTTYSNNITVTS--STLKSILTSVVS 937
QY 466 -ASVPGTCAATGAG-----TVSSVTVTSWPSIVATGTTTATPTPGSGSVTSTKTAT 520
DB 938 TASSYIASHASNSLTETPKTFSSSSTLG-ESLISINTNSLIVKPPSSLSLSSSTTSGLTSS 996

Qy 521 ASKSTTTRS 530
Db 997 SSTSPSTRS 1006

RESULT 33

JC2143

ice nucleation active protein - *Erwinia uredovora* (strain KUIN-3)C:Species: *Erwinia uredovora*

C:Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 17-Mar-1999

C:Accession: JC2143

R:Michigami, Y.; Abe, K.; Obata, H.; Arai, S.

Biosci. Biotechnol. Biochem. 58, 762-764, 1994

A:Title: Cloning and sequencing of an ice nucleation active gene of *Erwinia uredovora*.

A:Reference number: JC2143; PMID:94264407; PMID:7764866

A:Accession: JC2143

A:Molecule type: DNA

A:Residues: 1-1034 <MIC>

C:Comment: This protein consists of 52 repeats of closely related 16-amino acid motifs

C:Superfamily: ice nucleation protein

F:161-993/Region: R-domain

Query Match 5.5%; Score 151.5; DB 2; Length 1034;
Best Local Similarity 21.9%; Pred. No. 0.038;
Matches 127; Conservative 60; Mismatches 210; Indels 183; Gaps 25;

Qy 47 GADGAVSGADSGIVVASTPDNPDYFYWTRSGLVLTUPLFRN-----GDTSL 99
Db 299 GADSSIIAGVSTQTAGESTQTAGVSTQTAGK-----SDLTAGVSTGTAGDSS-- 350
Qy 100 STIENTISAQAIYQGISNPSG-----DLSSG-----AGIGEPKFN 134
Db 351 SLIAGVSTQTAGEDSLTAGVSTQTAGKSGDLTAGVSTGTAGADSSIIAGVSTQTA 410
Qy 135 VDETAVTGWGRPQRDGPALRATAMIGFCQWLLDNGYSTATDIWPLVENDLSYVAQYV 194
Db 411 GEESTQTAGVSTQ-----TAQKSD--LTAGVSTGT-----AGDDSLIA-- 450
Qy 195 NQCYGLMEVNVSSFFITIAVQHRALVEGSAFATAVGSSGSCWDSQAPELCY--LQSF 252
Db 451 ---GYSTQTAGEDSSLTAGVSTQTAGKSGDLTAGVSTGTAGVSSSLTAGVSTQTAG 507
Qy 253 TGSFIANFDSRSRGADANTLL---GSIHTFDEPAACDDSTFQPCSPRALNKEVDS- 308
Db 508 YGSLTLAGVSTQTAGNESPLITGYSTGT---AGANSLIAGVSTGTASTNSVLTAG 563
Qy 309 FRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFWFLCTLAABQLYDALYQWDKGSLVET 368
Db 564 YGSLTQTAGREG-SDULTA-----GYGSTQTA 586
Qy 369 DVSLEDFKALYSDAATGTYSSSSTYS--IYDAVKTFADGFVSIYETHAASNGSSEQY 426
Db 587 QEN-----SDLTGYGSTGTAGYDSSLIAGVSTGTAGYSHIL---TAGVSTQTAG 635
Qy 427 DKSD-----GEOLSR---DLTWSYAL 446
Db 636 ERKDLTTGYSTGTAGADSSLIAGVSTGTAGYNSILTAGVSTQTAGNESDLITGYST 655
Qy 447 LTANNRNSVVPASWGETSASVPGTCAA-----TSAIGYSSVTVTSW-PSIV 494
Db 696 STAGYESSLT--AGVSTQTAGFKSTLMAGYSSQRTAREQSSSLTAGVSTSMAGYDSSL 753
Qy 495 ATGCTTTT-----TPFGGSV-TSTSKTATATASKISTTT 528
Db 754 AGYGTQTAGYQSLTAGVSTQTAGHSTLTLAGYGSTAT 793

RESULT 34

S25370

MSB2 protein - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein G4017; protein YGR014w

C:Species: *Saccharomyces cerevisiae*

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000

C:Accession: S25370; S64305
R:Bender, A.; Pringle, J.R.
Yeast 8, 315-323, 1992
A:Title: A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence defect.

A:Reference number: S25370; PMID:92383951; PMID:1514328

A:Accession: S25370

A:Molecule type: DNA

A:Residues: 1-1306 <BEN>

A:Cross-references: GB:M77354; NID:g171993; PIDN:AAA4798.1; PID:g171994

R:Rieger, M.; Mueller-Aber, S.; Brueckner, M.; Schaefer, M.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64305

A:Accession: S64305

A:Molecule type: DNA

A:Residues: 1-1306 <RIP>

A:Cross-references: EMBL:Z72799; NID:g1322977; PID:g1322978; MIPS:YGR014w

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:MSB2

A:Cross-references: SGD:S0003246; MIPS:YGR014w

A:Map position: 7R

C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

C:Keywords: transmembrane protein

F:3-19/Domain: transmembrane #status predicted <TM1>

F:1189-1205/Domain: transmembrane #status predicted <TM2>

Query Match 5.5%; Score 151.5; DB 2; Length 1306;
Best Local Similarity 22.9%; Pred. No. 0.053;
Matches 134; Conservative 83; Mismatches 240; Indels 129; Gaps 24;

Qy 25 ATLDSWLSNBAVATARTAI-LNIGADGA---WVSGADSGIVVASTPDNPDYFYWTRD 79
Db 267 APLQTESSSFTTASALVPVSSSTDVDSASAPVSSAAGQIASSTSTNFTMSEFSLT 326
Qy 80 SGLVLTLYDLFRNGDTSLLSTIENYISAQAIYQGISNPSGDLSSGAGLGEPPFNVEDTA 139
Db 327 S-----TEVD---GSDVS--STVSALLSAPFLQTSNFSFIYSPVSPVSSQSSSD-- 373
Qy 140 YTGSMWRPQRDGPALRATAMIGFCQWLLDNGYSTATDIWPLVENDLSYVAQYV 199
Db 374 -----VASSTANVSSSPSDIPQSTSGSV-----SVAQSASALAF 412
Qy 200 DLWEVNGSS-----FETIAVQHRALVEGSAFATAVGSSGSCWCD-----SQAPEL 245
Db 413 QSTEVYGAASATMSGLSTSLQSTTL-DSSSLASSASASDLDYGVSTASTPLLS 471
Qy 246 CYLQSFMTGSFTL---ANFDSRSRGKD-ANTLLGSIHTFDEPAACDDSTFQPCSPRALA 300
Db 472 ASRQASTSSSFVSPSVSFVPSQSSDVASTSAPVVS-----SFSYTSIQAG 522
Qy 301 NKEVDSFRSIYTLNDGLSDSEAVAVGRY---PDTYNGN---PWFLCTLAABQL 352
Db 523 SSMWNPSSSTIYSSSTGSSSEEAATASATLSSGSSSTYAGNLQSPSTSLLESQA 582
Qy 353 -----YDALYQWDKGSLVETDVSLEDFKALYSDA---TGTYSSSS 391
Db 583 TSTSAVLASSSVTTSPTTAGCASTEASLLISSTSAETSVQSYSSQSTTALOTSSPASSS 642
Qy 392 -----STVSIYDAVTFADGFVSIYETHAASNGSSEYD-KSDGEOLASRD 438
Db 643 TTEGSEFTSQGFSTSLVLOMPSSISSEFSPQTTQNMASASSSQYTTSITLQVSD 702
Qy 439 LTVSYALLLTANNRNSV--VPASWGETSASVPGTCAATSAIGYSS-----VTV 487
Db 703 TSVSYT---TSSSSVQVSDTPVSY-TTSSSSVQVSDTPVSYTTSSSSVQVSDTPVSY 758
Qy 488 TSWPSIVATGCTTTTATPFGSGVSTSTSKTATATASKISTTTTSGMS 533
Db 759 TTSSSSVQVSDTPVSYTTSSSSVQVSDT---SVPTSSSSSVS 800

RESULT 35

A46954

mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum.
N/Alternate names: Beta-mannanase
C/Species: Caldocellum saccharolyticum
C/Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
A/Accession: A48954; MUID:93119139; PMID:1476429
R/Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A/Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A/Reference number: A48954; MUID:93119139; PMID:1476429
A/Accession: A48954
A/Status: preliminary
A/Molecule type: nucleic acid
A/Residues: 1-1331 <GIB>
A/Cross-references: GB:L01257; NID:g144290; PIDN:AAA72861.1; PID:g144291
A/Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBIPI:121577)
R/Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A/Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A/Reference number: A43745; MUID:91247819; PMID:2039230
A/Accession: B43745
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-337, 'PROHOHRO' <LUB>
A/Cross-references: EMBL:M6063; NID:g144292; PIDN:AAA72861.1; PID:g144294
A/Note: the authors translated the codon CAC for residue 262 as Glu
A/Note: this sequence has been revised in reference A48954
C/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 5.5%; Score 151.5; DB 2; Length 1331;
Best Local Similarity 21.1%; Pred. No. 0.054;
Matches 102; Conservative 60; Mismatches 161; Indels 161; Gaps 20;

QY 84 LKTLVDLFRNGDTSLSTIENYISAQIVOGISNPS-GDISGAGLGEPRNDETAYTG 142
DB 3 LKTKI--RKKWLSVLTCTVFLNLIIFIANVTLIPKGAATSDNGV---WKIDTSLIG 55
QY 143 S-----WGRPQRPALRAMTGFQW-----LIDNGYTSATDIWPLVANDLSYV 190
DB 56 TNHAHCYRRRLD-----TALRGIRSGMNSVRVLAISNGR----- 91
QY 191 AQWNQGYDLMEEVNGSSFFITAVQHRALVEGSAFATAVGSSGCSQAPLICYLOS 250
DB 92 -----WIKIPA-----SEVANIIISRS 109
QY 251 FWTGSFILANFDSRSGKANTLLGSIHTPDPEACDSTFGQSPRALNHKEVVDSEFR 310
DB 110 LGFKAILLEVHDTTGYGEDG-----AAC-----SLA---QAVEYWK 142
QY 311 SITTLNDGLSDSEAVAVGRYPEDTYVNGN--PWFICTLAAEQLYDALYQ-----W 359
DB 143 EIKSVLDGNDPFIYINIGNP--YGNMNYQMWVNDTKNAIKALRDGFGKTTIMVDAPNW 199
QY 360 DKQGSLEVTVDLSDFKCA-----LYSDAATGTYSSSSSTYSIVDAVKTAFDGFVSIY- 412
DB 200 GQWMSNTRMDAGSIMADPLRLNVFSIHMGYV---NTASKVEEYIKSFVDKGLPLVI 255
QY 413 --ETHAASNGSMSEQYDKSDGEQLSARDLTW-----SYAALLTANNRRNSVVPASGWT 464
DB 256 GERGHGTDDPDEALIVRAKQYKIGLFMSWCGNSSYGYLDVMVNMVNNPPTPWCQ- 314
QY 465 SASVPECTCAATSAIGYSSVYTWSPSIYATGTTTATPTGSGYSTISKTATASKT 524
DB 315 -----WYKINAIGTSTPTPTS-----TVPTPTPTPTPTVATATPTPTPTVST 360
QY 525 STTT 528
DB 361 PATS 364

RESULT 36
S25345
probable membrane protein YCR089w - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypotheical protein YCR1102

C/Species: Saccharomyces cerevisiae
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
A/Accession: S25345; S19504
R/Milson, C.; Grisanti, P.; Frontali, L.
Yeast 8, 569-575, 1992
A/Title: The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chromos
A/Reference number: S25345; MUID:92375594; PMID:1523889
A/Accession: S25345
A/Molecule type: DNA
A/Residues: 1-1609 <MIL>
A/Cross-references: GB:X59720; EMBL:S43845; NID:g1907116; PIDN:CAA42254.1; PID:g1907227
R/Frontali, L.; Grisanti, P.
submitted to the Protein Sequence Database, March 1992
A/Reference number: S19504
A/Accession: S19504
A/Molecule type: DNA
A/Residues: 1-1609 <FRO>
A/Cross-references: EMBL:X59720; NID:g1907116; PID:e264634; PID:g1907227; MIPS:YCR089w
C/Genetic: SGD:PTG2
A/Status: preliminary
A/Cross-references: SGD:S0000685; MIPS:YCR089w
A/Map position: 3R
C/Keywords: transmembrane protein
P;4-20/Domain: transmembrane #status predicted <TM1>
P;1592-1609/Domain: transmembrane #status predicted <TM2>

Query Match 5.4%; Score 150; DB 2; Length 1609;
Best Local Similarity 20.2%; Pred. No. 0.089;
Matches 124; Conservative 92; Mismatches 211; Indels 186; Gaps 32;

QY 61 VVASPTDNPDIYTYTTRSGVLVLTLDLFR---NGDTSLSTIENYISAQIVOGIS 116
DB 312 VLTSNSID-PSLETTSEYSSITQLSLNRSSEYVFTASIASSTFGTDSATSLDPLIS 370
QY 117 NPSGDISGAGLGEPRFN-----VDETAYTGS-----WGRPQRPALRAMTGFQ 164
DB 371 SVGSTASSFVGISTANFISQNSNYPESTAGSSQYQDMSSSL-PLSQTT----- 421
QY 165 WLNDN-----GYTSTARD-----IYW-PL-----YANDLSYVA 191
DB 422 WVVINTTNQGSVTSSTPSPAYVSTATKTVDGIVTEVTCPLQTKQAIGVSSISVAP 481
QY 192 QYWNQGYDLMEEVNGSSFFITAVQHRALVEGSAFATAVGSS--CWCSQAPLICYIQ 249
DB 482 QASFSFG--SSILSNSTTLAASNV-----PESTAGSSQYQDMSSSLP---LSQ 528
QY 250 SFWTGSFILANFDSRSGKANTLLGSIHT-----PDP 282
DB 529 TTV---VINTTNQGSVTSSTPSPAYVSTATKTVDGIVTEVTCPLQTKQAIGISS 584
QY 283 EAACDSTFGQSPRALA-NHKEVVD-SFRSIYTLNDGLSDSEAVAVGRYPEDTYVNG-- 318
DB 585 STIASOTGKPSILILGLSIQLDATFKGTETINTHLM-TESTSI---TEPTYFSGTS 640
QY 339 NPWFICT---LAAEQLYDALYQMDKQSLBVTVDLSDFKALYSDAATGYTSSS----- 390
DB 641 DSFYLTCTSEVNLASLSISYPN-FSSSEGSTATITNSVTF-----GSTSKPSTSVSN 692
QY 391 -----SSTYSIVDAVKTAFDGFVSIYETHAASNG---SMSSEQYDKSDGEQ----- 433
DB 693 PTEASQHVSSVNSLDFDTNSNTEITTAIVLSNHKTSNMDYSLITTLQTKSGKQTLVLST 752
QY 434 -----LSARDLTW-----SYAAL-----LTANNRRNSVVP 458
DB 753 VTTTVNGATTEYTWCPASSIATTSISXKTLVLTTEVCVSHSECPVTITSVATSTID 812
QY 459 ASWGETSASSVPECTCAATSAIG-TYSSVYTWSPSIYATGTTTATPTGSGYSTISKT 517
DB 813 LL--STSSSTVLSTVSEGAKNPAASEVINT--QVSATSEATSTSTOVASATATASB 868
QY 518 TATASKTSTTTS 530
DB 869 SSTTSQVSTASST 881

Db 228 AGTEW-----AGETNYNDAPF-----TMYEKIVLDYSTGKITYDQS 269
Qy 392 SYSSIVDAVKTPADGFVSIIVETHAANGSMBOYDK-----SDPEQLSARDL---T 440
Db 270 GSWESIE-----ADG-----GSYGNIDQAQEDPAVLANAGSISSSSSTSSST 311
Qy 441 WSYAALLTANNRRNSVVPASWGETSASSVPGTCATSAIGTYSV-----TTSWPSIVA 495
Db 312 VSSASSTVSSSVSTVSSASSSTVSSSSSVSSSSSSSTSPSSSTATSKTLAS 371
Qy 496 TGGTT-----TTAPYSGSVSTPSTKATATASKTS-----TTTSGM 532
Db 372 SSTTSSSISFEKQSSSSSKTVASSSTSESIISSTIKPATVSSSTRVAPTQSSV 431
Qy 533 S 533
Db 432 S 432

RESULT 39
glucan 1,4 alpha glucosidase (Glucosamylase) [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 07-Mar-2003
C:Accession: C90419
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.U.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.
arrett, R.A.; Ragan, M.A.; Seneen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: C90419
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-612 <KUR>
A:Cross-references: GB:AE006641; NID:G13815775; PIDN:AAK42610.1; GSPDB:GN00155
C:Gene: SSO2473
C:Superfamily: Sulfolobus type glucan 1,4 alpha glucosidase

Query Match 5.3%; Score 146; DB 2; Length 612;
Best Local Similarity 21.0%; Pred. No. 0.043;
Matches 101; Conservative 66; Mismatches 173; Indels 142; Gaps 23;

Qy 16 GLANVISKRATL-----DSWLSN-----EATVARTAIL--NNIGADGAWVSGA 56
Db 213 GLSDYVRKTPALIFKKVQYMWRLSKVNDYGEYDILRRSLILGSHVQNNCAIYASL 272
Qy 57 DSGIVVASPTNDNDYFTYTRDSGLVLTIVDLFRNGDTSLSLTENYISAQAIYVGIS 116
Db 273 DTIMKFNBDT---YVYVHRDAVFCILAL-----ELMGYPDRSRQFPEFTRLP 319
Qy 117 NPSGDSLSGAGLGEPRKFNVDETAATYTSWGRP-----ORDGPALATAM-IGFGQWL 167
Db 320 TINGALPH-----KTTVD--GHFGSTWHPWTLDIYPIQEDETALVYLMHFHFSKM-K 369
Qy 168 DNGYSTATDIWPLVENDLSYVAQYWNQ--TG-----YDLMEEVNGSSFITIAVOHRAV 221
Db 370 DVDPFIKT---YRPMVKGIDFLVYRREKATGLPSPFLMEERIGHFTTTIVINGLR 426
Qy 222 EGSAPFAIVASS--CSWCDQAPETILCYLQSFWTGSTILANFDSRSRGK-----DAN 271
Db 427 AADDFAYFGEDELAKQRYEVADQMRNSLDLFWVGDH-YARTIYMKGQVHKIDETVSS 485
Qy 272 TLGSIHTPDPEAACDSTPOCSPRALANHKEVVDSPRSIYTLNDGLDSEAVAVGRYP 331
Db 486 ILAAPTINVP--MSDSRF-----VKDLVTEKLSVKGGLV--RYE 523
Qy 332 EDYTY-----NGNPFCLTAAAEQYDALYQMDKQSLLEVTVDSLDFKALYSDAATGY 387
Db 524 GDQYLRGNNSNIMFISTLWLS-QVYSIMGEKDK----- 556
Qy 388 SSSSSTYSIVDAVKTPADGFVSIIVETHAANGSMSEQYDKSGEQLSARDLTWSYALL 447

Db 557 -----AAEKIDWVLSKSLPTGVLPQIDND-KYPSVPLAMSHBELI 598
Qy 448 TA 449
Db 599 RA 600

RESULT 40
F86719
hypoetical protein ynhg [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: F86719
R:Boletijn, A.; Wincker, P.; Mager, S.; Jaillon, O.; Malarme, K.; Weisenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi
A:Reference number: A86625; MUID:2125186; PMID:11337471
A:Accession: F86719
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-614 <STO>
A:Cross-references: GB:AE005176; PID:G12723675; PIDN:AAK04856.1; GSPDB:GN00146
C:Experimental source: strain IL1403
C:Gene: ynhg

Query Match 5.2%; Score 144.5; DB 2; Length 614;
Best Local Similarity 21.9%; Pred. No. 0.054;
Matches 138; Conservative 89; Mismatches 225; Indels 179; Gaps 34;

Qy 5 SLIATSLGV-CTGLANVISKRATLDSWLSN--EATVARTAILNNIGADGAWVSGADSGIV 61
Db 16 SLIATSLTAGFSGIATVNAKTTVADQVNGEVAIYAQSAAGNQNPQO-----NNGLV 68
Qy 62 VASPTNDNDYFTYTRDSGLVLTIVDLFRNGDTSLSLTENYISAQAIYVGISNPSGD 121
Db 69 --DSDGNP-----VNDSGIANNQ-----TGQAGNTSGTKNDGSOQTTLBS-STPS-- 111
Qy 122 LSSGAGLGEPRKFNVDETAATYAGSWGRFQPDGPALRATAMIGFGQWLNDNGYSTATDIWVP 181
Db 112 ---TMAVTFSA--TKYVG-----GVPRGYVDPSTDTTS--VAP 145
Qy 182 LVENDLSYVAQYWNQ--GYDLWEE--VNGSSE--FTIAVOHRAVLEG-----SA 225
Db 146 -VTTDASGLADFGTLDGYLTFHQVTTVNGITTVGDFIVOVSHEDSQAGIVNYPRLDMS 204
Qy 226 FATAVSSCSW-----CDQAPETILCYLQSFWTGSTILANFDS-----RSGKD 269
Db 205 SSAGLGTSATTNADNPNNGTPEQIANPNAVGNSDOTLTNTDNNAGNENLANGTWTNGSD 264
Qy 270 -----ANTLISGSIHT-FDPEAACDS-----TFQPCSPRALANHKEVVD 308
Db 265 NONTTAAAGNTVANNVNVTFDSSQTNNGNGTGVGYIVLDQ--LPNNLVNVSSTV--- 319
Qy 309 FRSIYTLNDGLDSEAVAVGRYPEDY---NGNPFCLTAAAEQYDALYQMDKQGS 364
Db 320 --TVSTVIYVTVNGSGTKVGTLPPTDYITNGNKKIVTLLTTAQQHAAASLGSADGA 377
Qy 365 LEV---TDVSLDFPKALYSDAATGYSSSSSTYSIVDAVK--TFADGFVSIIVETHAASNG 420
Db 378 LNIITPSTYKSAIGSA--TDSATTTTMAAGADLSTTAAKSTLVNGLGEMTTIDASTNA 435
Qy 421 SMS-----EQYDKSGEQL-----SARDLTWSYALLTANNRRNSVVP 458
Db 436 ALAGATFTVVRANKEDADQFVEANAAFPNNSASGGTIVNLTJSSKAAFTYGDTSGNA--- 492
Qy 459 ASWGETSASSVPGTCATSAIGTYSVTVTWSPIVATGCTTT-----ATPTG--- 507
Db 493 ---NTSATAPVFTTGKQGIATFNGLINLVD---NNTDGSNTNYLIVEAAPTYQLP 544
Qy 508 -----SGSVTSTSKTRATASKTSTTRS 530

Db 545 SVTTAANLGAATA---STAPATDTTITNN 572

RESULT 41

T11678

hypothetical protein SPBC21D10.06c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999

C:Accession: T11678

R:Seeger, K.; Harrie, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z17313

A:Accession: T11678

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-948 <SEE>

A:Cross-references: EMBL:AL031536; NID:e1319499; PID:e1319505

A:Experimental source: strain 972h(-)

C:Genetics:

A:Map position: IIR

A:Note: SPBC21D10.06c

Query Match 5.2%; Score 144; DB 2; Length 948;

Best Local Similarity 20.4%; Pred. No. 0.11;

Matches 114; Conservative 84; Mismatches 234; Indels 126; Gaps 22;

QY 25 ATLDSLSNEATVARTAILNNGADGAWVSGADSGIVASPTDNDPYFTWTRDGLV 84

Db 211 SLSLSFVITNVDSSTTSVINIYA-----STLESSLTNTVSPTESTFETKSSTSVPT 265

QY 85 KTLVDLFRNGDTSLSLTIENYISAQAVOGISNP-----SGDLSSGAGLGEPPVDETAAT 141

Db 266 QTI-----DSSGFTS-----STPVLSTSSSTSSSSQDSTIIDTPET 304

QY 142 --GSGWRPORDG-----ALPATMIGFGQMLDNGYTSATDVIWPLVRNDLSY 189

Db 305 IATSTLOPTTSPITTSAPLSGALPTTPSSLSSTVEVEYFTKITL-----DTSS 355

QY 190 VAQYMNQGYDL---WEVNGSFFITIAVQHRALVGSAR----- 226

Db 356 IVTY--STVEVLTETETTSSEISIIYNFSTPIGSSFPDGFKPINPTSPSLTSSYK 413

QY 227 ---ATAVSSCSCWCDAQEPIILCYLO-SFWTGSFIIANFDSRSGKADANTLGSHTPDP 282

Db 414 KIPSTLTPRSSKMITTTTSVSNNQSSFLITSTFSSYHSPRYVSSPLTS----- 467

QY 283 EAACDDSTPOCSPRALANHKEVDSFRSITYLNDGLDSEA---VAAGRYPEDTY-YN 337

Db 468 -----NPFSSISHSASSLPIPPSSYLSNVTLLHSSVQSSQSFVSV---PSTQSY 518

QY 338 GNPWFICTLAAEQLYDALYQMDKQSLVETVSLDFPKALYSDAATGYSSSSSTYSS 397

Db 519 TSSNFTPTTITISL-----SSFTTIVSSSFQYSSLSNVTTNAAQSSLSSSSN 568

QY 398 VDAVKTFADGFVIVETHAASNGMSEQYDKSDGEOLASADLTWSY-AALLTANRRNSV 456

Db 569 SSALTHISSIYSSGSSSALSSTIYSSINSSSVIYSSSSQLQSSSVITTTTSGSY 628

QY 457 VPASWETSASVVP-GTCAATSAIGTVSSVTVTSWPSIVATGG---TTTATPTGS-G 509

Db 629 -----GFTTITATPVGSTAGTVV---DIPPSWTEETVTSVGFTTIATPVSTAG 679

QY 510 SVTSTSKTATASKSTT 527

Db 680 TVLVDPITPSPWTEIVTS 697

QY 510 SVTSTSKTATASKSTT 527

Db 680 TVLVDPITPSPWTEIVTS 697

QY 510 SVTSTSKTATASKSTT 527

Db 680 TVLVDPITPSPWTEIVTS 697

QY 510 SVTSTSKTATASKSTT 527

Db 680 TVLVDPITPSPWTEIVTS 697

QY 510 SVTSTSKTATASKSTT 527

Db 680 TVLVDPITPSPWTEIVTS 697

QY 510 SVTSTSKTATASKSTT 527

Db 680 TVLVDPITPSPWTEIVTS 697

R:Gaur, N.K.; Klotz, S.A.

Infect. Immun. 65, 5289-5294, 1997

A:Title: Expression, cloning, and characterization of a Candida albicans gene, ALA1, that

A:Reference number: Z20847; NMID:98053977; PMID:9393828

A:Accession: T30531

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1419 <GAU>

A:Cross-references: EMBL:AF025429; NID:g2522218; PID:g2522219; PIDN:AA08883.1

C:Genetics:

A:gene: ALA1

C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase }

Query Match 5.1%; Score 142; DB 2; Length 1419;

Best Local Similarity 19.9%; Pred. No. 0.26;

Matches 134; Conservative 89; Mismatches 210; Indels 242; Gaps 34;

QY 26 TLDSWLSNEATVARTAILNNGADGAWVSGADSGIVASPTDNDPYFTWTRDGLV 83

Db 440 TQOFW--SESFSTTTITNSL-----KGTDS-VIVREP--HNPTVTTEFWS-ESFAT 486

QY 84 LKTLVD-----LFRNGDTSLSLTIENYISAQAVOGISN-PSGDLSSGAGLGEPP 131

Db 487 TETITSKPEGTDVIVREPHNPVTITTEFWSSEYATTEITNGPEG--TDSVIVREPHNP 544

QY 132 -----KFVDETAATVGS-WGRPORD-----GPALPATMIGFGQMLDNGYTSAT 176

Db 545 YTTTTFKFSSEYATTEITTNKEGTDVIVKEPHNPVTITTE-----W-SSSYATTE 597

QY 177 DIWVP-----LVR--NDLSVAQYMNQGYDLWEVNGSFFITIAVQHRALVGSARA 227

Db 598 ITNGPEGTDVIVREPHNPVTITTEFWSSES-YATTEITNGPEGTDVIVHDLSSSSST 656

QY 228 TAVSSCSCWCDAQEPIILCYLOSFWTGSFIIANFDSRSGKADANTLGSHTPDPACD 287

Db 657 TAIBESSDNISSAOR-----SSSVEQSSSIYGLSSSDIPLSSG 697

QY 288 DSTPOCSPRALANHKEVDSFRSITYLNDGLSDSEAVAVGRYPEDTYNNGNPWFICTLA 347

Db 698 ---NPSS-----STGLTSSSSSTVSDSD--SSSIESSTLS 731

QY 348 AAEQ---LYDALYQMDKQGS-LEVTDV---SLD----- 373

Db 732 SDRCCSSISDPTSPFDSSSDLESTITWSSSIDQSHLVQSVNSISTQELSSSS 791

QY 374 -----FPKALYSDAATGYSSSSSTY-----SSIYDAVK 402

Db 792 EESSTPATDALVSSDASISLSDTSYPSSTISSSDDEPHTIAGSDLSISFITSTVE 851

QY 403 TRADGFVSI VETHAAS-NGSMSEQYDKSDGEOLASADLTWSYAALLTANRRNSVVP--- 458

Db 852 ISSDS-VSLTSPDASSFDSSSLNSDSSSPSSDQSD-----ILTSSPSTLVVPSFS 903

QY 459 -----ASWGETSASVPGTCA-----ATSAIGTY 482

Db 904 LSSSSLSLITYHYNSTTYHASESSSSVAPSMASEANDDTYLLSSTDTTSSIGTD 963

QY 483 SSVT-----VTSWPS-----IVATGTTTATPTGS-GSVTSTSKTTA 519

Db 964 SSTVTFCRDNGDGCIVTGMPSISIDSEQTSVTVTTSSFVASTPTPSAEOISITDNPIDS 1023

QY 520 TASKSTTRSGMSL 534

Db 1024 SOTSASSSTKSSVSV 1038

QY 520 TASKSTTRSGMSL 534

Db 1024 SOTSASSSTKSSVSV 1038

QY 520 TASKSTTRSGMSL 534

Db 1024 SOTSASSSTKSSVSV 1038

QY 520 TASKSTTRSGMSL 534

Db 1024 SOTSASSSTKSSVSV 1038

QY 520 TASKSTTRSGMSL 534

Db 1024 SOTSASSSTKSSVSV 1038

QY 520 TASKSTTRSGMSL 534

Db 1024 SOTSASSSTKSSVSV 1038

R, Millward-Sadler, S.J.; Poole, D.M.; Henriessat, B.; Hazlewood, G.P.; Clarke, J.H.; Gilh
Mol. Microbiol. 11, 375-382, 1994
A:Title: Evidence for a general role for high-affinity non-catalytic cellulose binding
A:Reference number: 140712; MUID:9422415; PMID:8170399
A:Accession: 140712
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-644 <RSS>
A:Cross-references: EMBL:X76729; NID:g558176; PIDN:CA54145.1; PID:g558177
C:Genetics:
A:Gene: xymD
C:Function:
A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic bonds in xylans
A:Pathway: xylan degradation
C:Superfamily: Cellulomonas endo-1,4-beta-xylanase D; endo-1,4-beta-xylanase homology;
C:Keywords: extracellular protein; glycosidase; hydrolyase; polysaccharide degradation
F:1-35/Domain: signal sequence #status predicted <SIG>
F:36-644/Product: endo-1,4-beta-xylanase D #status predicted <MAT>
F:361-508/Domain: nodB homology <NODB>
F:126-216/Active site: Glu #status predicted

Query Match 5.1%; Score 141.5; DB 1; Length 644;
Best Local Similarity 20.7%; Pred. No. 0.093; Matches 142; Conservative 76; Mismatches 238; Indels 229; Gaps 35;
QY 28 DSWLSNEATVARTAI--LNNIGADGAVSGADSGIVASPS---TDN-----PDYFYT- 75
DB 3 DSEBATTTTTRRRRLQALTGILMAGALVACA---LAASPPAAAVTNTTGTGHDGYYSF 59
QY 76 WTRDSGLVLTVDLFRNGDTSLSTIENTYISAQIVQGISNPSGDISGAGLGEPPFNV 135
DB 60 WTPSGSV---SMDLNSGGGYTRWSNTGNFVAG---KGMSTGGRKTVSYSG---QFNP 108
QY 136 DEFAYTSGKRPQRDGPALATAMIGCOMLDNGYSTATD-----IWPVLV 183
DB 109 SRNAYLTLYGMTQ--SPLVEYIVDSWGYRPTGTFTMGITVSDGGIYDRTORVNFPSI 166
QY 184 RNDLSYVAQYWN---QTG-----YDLME-----EYNGSSPF 211
DB 167 EGSSSTFYQWYSVRQCRGTGTTISGNHFDAMAKGNLGRHNYMIMATGYSOSSGSSI 226
QY 212 TIAVQHRALVEGSAFA---TAVGSSCSWCDSPAPELLCYQSF-----WTGSF 256
DB 227 TVS-----EGSGCGGCGDTGCGGSGTCSVTATRAEWSDRPNVTYVSAGSAMTVNL 279
QY 257 ILANFDSRSGKQAN-----TLGSIHF-----DEAACDOSTPQCS 295
DB 280 ALNGSQTIIQASWNNVNTGSGSTRVTVPNGSGNTGVTVMKNGSSTTPAATCAGSGGCTAT 339
QY 296 PRALANKEVVDSPRSY---TLNDG-----LSDSEAVAVGRYPEDTYNGNP 340
DB 340 PTPPTPTTPPGSSAGYVGLTFDDGPNGTGTTNIIISTLTQYGAFAI--VPTTGNAGGNP 398
QY 341 WFLCTLAABQLYD---ALYQWD-----KQSLLEVTDVSLDFP 375
DB 399 SLM-----QAYKNAGVQIGNHSDHPLVNMSCDMOSQLTRTQOAIQDTAGVPTLFI 451
QY 376 KALY--SDAATGYSSS-----SSTYSYDAKTPADGVSLVE--- 413
DB 452 RPPGESNATLRQVESSLGLREIIVDVSODMNNASASQIRQAASRLTNQIILMHMPA 511
QY 414 -THAASGMSSEQYDKSDGEOLASRDLTWSYALILTANNRNSVVPAS--WGTSASSVPG 471
DB 512 ATVOALTGIL-----QDLRSRLCTGHISSTG---RAVAPSSAGGGGGGGGCTG 558
QY 472 TCAATSAIG-----TY-----SSVTY-----SWPSIVANGTTTATPT 506
DB 559 SCVSASVARGEMADRFNVTYVSQSSSWVYTLGLNGSGVQSSWNAALTGSSGTVTAPN 618
QY 507 GSGS---VT--STKTTATASKTSTT 527
DB 619 GSGNSFGVTYFKNGSSATPGATCAT 643

RESULT 44
B86807
hypotheical protein yoiC [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C/Species: lactococcus lactis subsp. lactis
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #ext_change 03-Aug-2001
C/Accession: B86807
R/Biologic: A.; Wincker, P.; Manger, S.; Jaillon, O.; Maizarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: B86807
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1441 <STO>
A:Cross-references: GB:AE005176; PID:g12724450; PIDN:AAK0556.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yoiC

Query Match 5.1%; Score 141.5; DB 2; Length 1441;
Best Local Similarity 19.3%; Pred. No. 0.29; Matches 119; Conservative 87; Mismatches 237; Indels 175; Gaps 24;

QY 13 VCTGLANVISKRATLDSWLSNEATVARTAILNIG-----ADGA--WVSG 55
DB 438 ITQGISNVIOE-----IANNIGNALNGNIIVSSLIITIANGLQTVTA 479
QY 56 ADGGIVASSTNDPNDFYTWTRDS--GLVLTVDLFRNGDLSLSTIENTYISAQ--- 109
DB 480 VTNGLI---PTDVIONFTITNVNSAFGSTNISLNLGLSGTSTVTNALSITGAOLPPI 535
QY 110 -----AIVOGISNPSGDISGAGLGEPPFNVDETYTSGWGRPQRDGPALRA 156
DB 536 NQSIGLVGNVSGITQGIOTGQVVS-----LINTLATLNGVAANVQASQDQMFA 588
QY 157 TA-----MIG---FQGMIL-----DNGYSTATDIWPL--VRNDLSYVAQYWNQT- 197
DB 589 LAKEPKEVINANGTFNESFLANTONADGTYVSDTGSHSPLTALTNKVALVYNGYIQAT 648
QY 198 -----GYLMEEVNSSFFTTAVQHRALVEGSAFAFTAVAGSSGSCWCDSPAPELLCY 247
DB 649 SVGLKTIIVTIPQMLIDAIAGLSNPNLGLNTGSLDG--VENTVINGVNVGVSFASAMQG 707
QY 248 LOSFWTGSFLIANFDSRSGKDNATLLGSIHTFDPEAACDSTFQCSPRALANKEVVD 307
DB 708 LNSVTAGASDLVUNTSNWTGTSNIVSQY-----VNSIAN-----A 744
QY 308 SFRSIYTLNDGLS---DSEAVAVGRYPEDTYNGNPWFLCTLAABQLYDALYQMDKQGS 364
DB 745 SATGVAIVTVGTPAADPSQIATNLATANHYYOGO-----SAATQROLDNVFPPE---T 796
QY 365 LEVTVDSLDPFKALYSDAALGTYSSSSSSTYSIYDA--VKTFPDGFYSIETHAASGSM 422
DB 797 FKVTPTPAL-----AGQNTTDTYSVVQYQNTNTIKGSIYVPDSTKALANAK 844
QY 423 SEQYDKSDGEOLASRDLTWSYALILTANNRNSVVPRA----- 459
DB 845 ATAVERKSDETVAPAPAPAGY--VLANDVNKAQVADANGVINFYDALIOPTLTKOSTID 903
QY 460 ---SMGTSASSVPGTCAATSAIGYSSVTYTSWPSIVATGTTTAT---PTGSGSVTS 513
DB 904 EGSITW--TAADNFTGGSSTGTIVTINDVTGTVDTSTPGVSSVYTYDPTGASISS 961
QY 514 TSKTTATASKTSTTTBSG 531
DB 962 VANITVDSSTNTNPTDG 979

RESULT 45
S11672
ice nucleation protein - Xanthomonas campestris

THIS PAGE LEFT BLANK